AD-A119 049 PENNSYLVANIA STATE UNIV UNIVERSITY PARK APPLIED RESE--ETC F/G 17/2
EVALUATION OF CRITICAL BANDWIDTH USING DIGITALLY PROCESSED SPEE--ETC(U)
MAY 82 R D CELMER N00024-79-C-6043
UNCLASSIFIED ARL/PSU/TM-82-114 NL

1 OF 2
AD A
119049

AD A119049

EVALUATION OF CRITICAL BANDWIDTH USING DIGITALLY PROCESSED SPEECH

Robert D. Celmer

Technical Memorandum
File No. TM 82-114
May 12, 1982
Contract No. N00024-79-C-6043

Copy 5

The Pennsylvania State University
Intercollege Research Programs and Facilities
APPLIED RESEARCH LABORATORY
Post Office Box 30
State College, PA 16801

APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED

NAVY DEPARTMENT

NAVAL SEA SYSTEMS COMMAND

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER TM 82-114	2. GOVT ACCESSION NO. AD-A119 049	3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle) EVALUATION OF CRITICAL BANDWIDTH USING DIGITALLY PROCESSED SPEECH	5. TYPE OF REPORT & PERIOD COVERED Ph.D. Thesis, August 1982 6. PERFORMING ORG. REPORT NUMBER TM 82-114
7. AUTHOR(s) Robert D. Celmer	8. CONTRACT OR GRANT NUMBER(s) N00024-79-C-6043
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Pennsylvania State University Applied Research Laboratory, P. O. Box 30 State College, PA 16801	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Sea Systems Command Department of the Navy Washington, DC 20362	12. REPORT DATE May 12, 1982 13. NUMBER OF PAGES 106
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified, Unlimited 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited, per NSSC (Naval Sea Systems Command), June 18, 1982.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

thesis, digital, processing, speech, frequency, bandwidth

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Existing literature suggests that the hearing mechanism deals with incoming speech material by filtering the signals into a series of frequency bands. The width of these bands has been referred to as the critical band, that is, the perceptual frequency bandwidth observed in a variety of psychoacoustic contexts. Digital processing techniques have been developed for altering available recorded speech materials so that the frequency resolution available in the resultant stimuli may be controlled. Tapes have been produced wherein the frequency bandwidth resolution is limited to no better

than one critical band and these tapes have been used in intelligibility testing. Some existing research indicates that the critical band is significantly widened in many individuals with sensorineural hearing loss of cochlear etiology. The digital processing routines described above were also used in developing tape recorded materials with bandwidth resolution limits considerably wider than the normal critical band. The bandwidths chosen for this stage of the digital processing were based on empirical observations of the critical band of sensorineural hearing impaired patients. These recordings were also used in intelligibility testing with normal listeners. Implications of these studies for the clinical measurement of speech intelligiblity will be discussed.

ABSTRACT

Existing literature suggests that the hearing mechanism deals with incoming speech material by filtering the signals into a series of frequency bands. The width of these bands has been referred to as the critical band, that is, the perceptual frequency bandwidth observed in a variety of psychoacoustic contexts. Digital processing techniques have been developed for altering available recorded speech materials so that the frequency resolution available in the resultant stimuli may be controlled. Tapes have been produced wherein the frequency bandwidth resolution is limited to no better than one critical band and these tapes have been used in intelligibility testing. Some existing research indicates that the critical band is significantly widened in many individuals with sensorineural hearing loss of cochlear etiology. The digital processing routines described above were also used in developing tape recorded materials with bandwidth resolution limits considerably wider than the normal critical band. The bandwidths chosen for this stage of the digital processing were based on empirical observations of the critical band of sensorineural hearing impaired patients. These recordings were also used in intelligibility testing with normal listeners. The critical bandwidth of both normal and sensorineural hearing impaired listeners has been measured by an independent technique. Tapes have been produced wherein a complex of tones vary with time from a sub-critical to a supra-critical bandwidth. The bandwidth at which a perceptual change in the test signal occurred was recorded as that listener's critical bandwidth. The results of this independent critical bandwidth test were found to be correlated to the results of the digitally processed speech discrimination test. Implications of these studies for the clinical measurement of speech intelligibility are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

The author wishes to express gratitude to his teachers, especially Dr. Gordon R. Bienvenue, Dr. Paul L. Michael, Dr. James Martin and Professor James Prout.

The author is also indebted to the generous assistance provided by Mr. Daniel Richards, Ms. Fran Murphy, Ms. Linda Bonsall, Jane Keyes and Cindy Klobuka.

A gracious "thank you" is extended to Professor Conrad Hemond for his encouragement to pursue graduate studies.

Heart-felt appreciation is extended to Dr. Walter D. Celmer, whose love and encouragement has always been felt.

Virginia S. L. Celmer: As the backbone of this graduate effort, the author is eternally grateful.

This research was supported in part by the Applied Research Laboratory at The Pennsylvania State University under contract with the Naval Sea Systems Command.

Chapter I

BACKGROUND

Brief Physiological Overview

Detailed descriptions of auditory anatomy may be found in such sources as Fex (1962), Goldstein (1968), Milner (1970), Dallos (1973) and Spoendlin (1973). The portion of the inner ear primarily responsible for the hearing process is called the cochlea. This coiled, membranous structure contains three compartments, or scalae, separated lengthwi along the human cochlea's two-and-three-quarter turns. The scala media houses the Organ of Corti, whose complex set o receptor cells serve as the locality for conversion of continuous acoustic waves into neural impulses. An acoustic input arriving at the oval window via the ossicles gives rise to transverse and longitudinal waves in the cochlea's endo- and perilymphatic fluids, respectively. Wave propagation in the cochlea is essentially a case of wave propagation in a shallow fluid of nonuniform depth (i.e., similar to ocean waves approaching a beach). The scala tympani and the scala vestibuli become gradually "shallower" towards the apical end. The point at which the transverse wave at the fluid interface of these two scalae (i.e., the scala media duct) will crest is dependent on its frequency (Dallos, 1973). High frequency waves crest near the basal end (far from "shore") while low frequency waves consequently crest close to the apical end. Maximal dissipation of the wave's energy occurs at this location. The inner and outer hair cells of the Organ of Corti are embedded in the basilar membrane. These hair cells have cilia (hair-like projections) which are rooted in the surface plate and the ends of these cilia either float freely in the endolymph or course into the opposite tectorial membrane. Wave motion causes the cilia to undergo shear, inducing the corresponding hair cell to trigger the neuronal fibers synapsed at its base. It should be noted that the hair cells existing at the wave's crest point will undergo maximal shear and thus relay the most neural information. Thus, the hydrodynamic construction of the cochlea, combined with the morphology of the Organ of Corti, yields a preliminary place-specific frequency analysis of the acoustic signal.

Bipolar cells synapse at the base of the hair cells with neurons that carry frequency, phase, and amplitude information towards the brain stem. The collection of the cell bodies of these neurons is called the spiral ganglion, located in that portion of the modiolus called the Rosenthal canal. The subsequent collection of the axons proceeding to the brainstem make up the bulk of the auditory nerve (VIII).

The human auditory nerve has been observed to contain about 31,000 neural fibers (cf. Gelfand, 1981). The dendrites of VIII enter the Organ of Corti at the junction of the lamina and the basilar membrane through openings in the lamina called the habenula perforata. Myelination is in evidence on that portion of the fibers exterior to the Organ of Corti, but not on the portion within.

The 31,000 fibers innervate the inner and outer hair cells in vastly different patterns and percentages. The inner hair cells receive about ninety-five percent of the fibers; each cell is innervated by about twenty fibers. This innervation occurs in a primarily radial fashion, or perpendicular to the turns of the cochlea. This divergent relationship between inner hair cells and the internal spiral bundle continues from base to apex.

The afferent neural supply for the outer hair cells is comprised of the remaining five percent of VIII. These 2,500 - 3,000 fibers cross between the pillar cells, pass through the tunnel of Corti, up between Deiter's cells and then spiral as the outer spiral bundle. Individual afferent collaterals each synapse about ten outer hair cells in a longitudenal (spiral) fashion, crossing the outer rows in a basalwards direction. This convergent innervation of the outer hair cells is found from base to apex.

Differences between the outer and inner hair cells also exist with regards to their efferent innervation. About 3,000 efferent fibers enter the cochlea at the same afferent locations, i.e., the habenula perforata. About three-quarters of these fibers originate at the contralateral superior olive and the other one-fourth at the homolateral superior olive. Once inside the cochlea, about eighty percent of these efferent fibers cross to the outer hair cell region via the tunnel of Corti. By this point, extensive ramification of the fibers has taken place, resulting in about 40,000 nerve endings which innervate the outer hair cell bodies themselves (presynaptically) in a radial fashion. The most extensive efferent innervation occurs on the basal turn, where each outer hair cell receives six-to-ten efferent endings, as opposed to the five-to-eight per cell ratio found towards the apical end. Inner hair cells receive efferent innervation from the remaining twenty percent of the efferent fibers in a different fashion. Collaterals from these crossed and uncrossed olivo-cochlear bundle fibers synapse on the afferent fibers of the inner hair cells, rather than on the cell bodies themselves as noted for the outer hair cells. Each inner hair cell's many afferent fibers receives an efferent nerve ending, with each collateral originating from a separate fiber.

Afferent Central Auditory Pathways. The axons of the bipolar cells described above make up the bulk of VIII. Proceeding towards the brainstem, fibers of VIII that originate within the apical turns of the Organ of Corti synapse on the ventral portion of the dorsal cochlear nucleus (DCN). Fibers from the basal portion of the Organ of Corti terminate on the dorsal portion of the DCN and the ventral cochlear nucleus (VCN). Second-order neurons relay information from these cochlear nuclei to either the superior olivary complex (SOC) or to the inferior colliculus (IC). Specifically, some fibers from the DCN terminate on the cerebellum, while most decussate to the contralateral nucleus of the lateral lemniscus (LL). Second-order posterior VCN fibers decussate to the contralateral LL tract and terminate on the dorsal-lateral portion of the IC. Anterior VCN fibers terminate on the homo- and contralateral medial accessory nucleus of SOC. Thus, each SOC would receive information from both cochleas, since the innervation patterns are mirror images of one another (cf. Gelfand, 1981). Third-order neurons arise from both medial accessory nuclei which ascend homolaterally within the LL, some of which terminate at the nucleus of LL, while most continue to the IC. Fiber tracts have been found between the contralateral nucleus of LL and the medially located reticular activating system (Ferraro and Minckler, 1977). At the level of the IC, decussation takes place between these two mid-brain structures via the commissure of the IC. Information is then transmitted almost entirely in a homolateral fashion via the brachium of the IC to their respective medial geniculate bodies of the thalamus. No further decussations take place as auditory radiations on each side each make a final, homolateral relay to temporal cortex. The first stop for an afferent message arriving at cortex is Broadman areas 41 and 42 (auditory area A-I of Woolsey and Walzl, 1942) of the temporal lobe, located anterior to the calcarine fissure.

Efferent Central Auditory Pathways. The existence of an efferent auditory pathway was discovered near the turn of the century by Ramon Y Cajal (1896) and elucidated by

Lorente de No (1937). Originating in the posterior cephalad of the diencephelon, the downward coursing fibers experience much the same afferent relays in the opposite order (Fex, 1962). The fibers follow a three-section route towards the SOC. First, the dorsal acoustic stria projects from the diencephalic area mentioned above in a medialwards direction over the restiform body to the medial geniculate. Next, the intermediate stria originate in the medial geniculate and decend ventrally and laterally to synapse at the "S" shaped principle olivary complex. Finally, Monakow's bundle originates at the IC and dorsal nucleus of LL and terminates at the DCN. It follows a pathway very similar to that of the intermediate stria's (Rasmussen, 1960).

Rasmussen (1946) made a highly detailed account of that portion of the efferent pathway progressing from the superior olives to the cochlea, naming it the olivo-cochlear bundle. The crossed olivo-cochlear bundle (COCB) originates in the contralateral medial SOC, crosses under the floor of the fourth ventricle, emerges from the medulla between the vestibular and pars intermedia nerves and continues laterally while running alongside the trigeminal nerve (V). Here these fibers turn slightly caudal, scatter around the vestibular nerve, reconverge near the spiral ganglion and then spiral in such a manner so as to distribute themselves to all turns of the Organ of Corti. Subsequent research by Fernandez (1951) noted that the branched fibers innervate a diffuse group of inner and outer hair cells. The uncrossed, homolateral component of the bundle originates in the ipsilateral SOC and joins the COCB lateral to the outgoing facial nerve. The homolateral component contains about one-fourth of the total fibers in the bundle; the COCB contains the remaining three-quarters (Fernandez, 1951).

The Critical Band Phenomenon

Existing literature on audition theory suggests a heavy reliance upon the notion of critical bands (e.g., Scharf, 1970). Most simply, a critical band may be conceived as an internal bandpass filter. In the initial conception of critical bands (Fletcher, 1940), the auditory system was theorized as consisting of a fixed bank of about twenty-four critical bands laid end to end, covering the audible frequency range. By contrast, current notions view critical bands as variable filter elements centered upon a particular signal frequency (cf. Scharf, 1970). The critical band has been defined empirically as "that bandwidth at which subjective responses rather abruptly change" (cf. Scharf, 1970, p. 159). In general, two stimuli separated in frequency by less than a critical bandwidth will interact in one of a number of ways, while two stimuli separted by more than a critical bandwidth will not. Changes of listener response due to the critical band phenomenon have been observed in such perceptual phenomena as masking, loudness, and musical consonance.

Masking and the Critical Band. A condition of masking is said to be in effect when a temporary loss of sensitivity to a stimulus occurs, caused by a simultaneous, ipsilateral presentation of another stimulus (Moore, 1977). By and large, a masking stimulus is most effective in hiding a given signal whenever the frequency content of the two signals is similar (Scharf, 1970). Under this context, then, a critical band may be defined as that masker frequency region wherein the masking of a given stimulus is most effective (Fletcher, 1940). Energy concentrations in regions larger than the critical band of the test signal demonstrate less efficient masking ability. A masker's efficiency may be defined as that amount of masking energy needed to provide a given threshold of masking. In a psychoacoustic context, a masking source is most efficient when its bandwidth lies within the critical band. Thus, the critical bandwidth mechanism seems to provide a resolution capability to the auditory system whose limit is reflected in its ability to mask signals.

Loudness and the Critical Band. Concomitant to the limits of masking bandwidth is the way a listener perceives the intensity of: 1) noise bands as a function of frequency bandwidth (Zwicker and Feldtkeller, 1955; Zwicker, Flottorp, and Stevens, 1957; and Zwicker, 1958) or 2) complex tones whose frequency separation is varied (Scharf, 1961, 1970). These studies showed a significant relationship between the bandwidth (Δf) of the stimulus and the point at which a listener hears a change in loudness:

> ...the loudness of a subcritical complex sound of invariant intensity is largely independent of Δf--it is about as loud as an equally intense pure tone lying at the band's center frequency. Only when Δf exceeds the critical band does the loudness of the complex begin to increase (Scharf, 1970, p. 161).

The process of integrating incoming stimuli for loudness perception, then, appears to operate through the filtering effect of the critical bandwidth mechanism.

Musical Consonance. Esthetic tests which rate the pleasantness of two-tone complexes have provided another setting under which the critical band phenomenon may be observed. Listeners were asked to rate the consonance of a pair of tones on a seven-point "pleasantness" scale. The overriding judgments of consonance were found at tonal separations of more than a critical band. A rapid decrease in the rating occurred as the tones moved to a separation narrower than a critical band (Plomp, 1964; Greenwood, 1961). These findings psychoacoustically demonstrate a phenomenon musicians have understood for centuries; consonance peaks occur at the common harmonic ratios--the fourth, fifth and octave (Plomp and Levelt, 1965). The relationship of these ratios to the critical band phenomenon in particular, further demonstrates the universality of this effect in hearing.

<u>Critical Bands and Speech</u>. Speech is the most pervasive and important acoustic stimulus for the human listener. Evidence suggests that the critical band may serve in the analysis of speech (cf. Scharf, 1970). The specific task of attempting to measure the critical bandwidth used in the analysis of speech sounds by the auditory system has been indirectly approached in the work of French and Steinberg (1947). In this study, subjects were asked to identify speech presented at different bandpass center frequencies and bandwidth conditions. The relative frequency components of speech were normalized a priori, taking into consideration the relative contribution of different frequencies to perceptual cues. The subjects' scores reflected their accuracy of perception as a function of total speech content. The results demonstrated a series of twenty-four bandwidths for which equal contribution to the perception of speech was realized. The twenty-four critical bands found in pure tone psychoacoustic studies were found to match very closely to the twenty-four bands which contributed equally to speech intelligibility. A verification study of French and Steinberg's work was conducted by Richards and Archbald (1956), who used only twenty variable passbands of speech with the Articulation Index. Equal contribution for speech was found for those bandwidths nearly equal to those of French and Steinberg.

Kryter (1960) conducted a baseline study of speech discrimination ability using bandpassed word lists. Center frequency, bandwidth and signal-to-noise ratio were varied. Although no subcritical bandwidths were presented, the results indicated that high intelligibility scores were possible with speech filtered to bandwidths equal to or wider than a critical bandwidth. A speech discrimination study using single passbands was conducted by Castle (1964), who varied the bandwidths and center frequencies of these signals. Although critical band data are not stated explicitly, the results indicated high speech discrimination scores for bands with widths greater than or equal to the critical bandwidths given in Table 1, while the scores dropped off at a significant rate for speech filtered through successively narrower subcritical bands. Chari (1977) presented listeners with single passbands of speech centered between 500 and 3150 Hz. The bandwidths were varied between a critical band and a one-third octave bandwidth. Good agreement with French and Steinberg's intelligibility scores was found using the critical bandwidth passbands. These research studies, then, have indicated the probable involvement of critical bands in the performance of speech listening.

In summarizing other research on critical bands, one finds two functional aspects of critical bands which appear to play an important role in the perception of speech, and hence, represent fundamental aspects of a listener's performance. As implied in the works of Fletcher (1940), Zwicker (1958), and Greenwood (1961), the critical band

Table 1

Examples of Critical Bandwidth

Number	Center Frequency (Hz)	Critical Band (Hz)	Lower Cutoff Frequency (Hz)	Upper Cutoff Frequency (Hz)
1	50	--	--	100
2	150	100	100	200
3	250	100	200	300
4	350	100	300	400
5	450	110	400	510
6	570	120	510	630
7	700	140	630	770
8	840	150	770	920
9	1,000	160	920	1,080
10	1,170	190	1,080	1,270
11	1,370	210	1,270	1,480
12	1,600	240	1,480	1,720
13	1,850	280	1,720	2,000
14	2,150	320	2,000	2,320
15	2,500	380	2,320	2,700
16	2,900	450	2,700	3,150
17	3,400	550	3,150	3,700
18	4,000	700	3,700	4,400
19	4,800	900	4,400	5,300
20	5,800	1,100	5,300	6,400
21	7,000	1,130	6,400	7,700
22	8,500	1,800	7,700	9,500
23	10,500	2,500	9,500	12,000
24	13,500	3,500	12,000	15,500

serves to band-limit background noise. The narrower the passband of the ear as a filter, the more noise the ear can reject, making it more tolerant to lower signal-to-noise ratios. Thus, a listener may be able to correctly perceive a spoken communication despite background noise simply because much of the energy associated with the noise lies outside the critical bands surrounding the formant frequencies of the speech.

Secondly, our ability to discriminate the harmonic content of complex signals (one of the many cues used, for instance, in speaker identification) is similarly related directly to the critical band phenomenon. Plomp (1964) and Haggard (1974) have demonstrated that listeners are able to discriminate those partials of a complex tone which lie more than a critical band apart. Morton and Carpenter (1963) found that formants can be identified by listeners even when no prominent energy peak is present as long as the most intense harmonics associated with each formant are separated by at least a critical bandwidth. Synthetic vowels presented to listeners by Remez (1977) showed an abrupt changeover from speech-like to non-speech-like sounds as the formant bandwidth increased to greater than a critical bandwidth. Preliminary analysis with reference to the critical bandwidth phenomenon indicates that this mechanism seems to be working for speech analysis on a peripheral basis.

Two functional characteristics of critical bands have been briefly reviewed (noise band limiting and harmonic discrimination). It is argued from these effects that critical bands play an important role in the correct perception of complex acoustic stimuli, such as speech.

<u>A Mechanism for Critical Bands</u>. A selective inhibition of frequency-specific afferent auditory messages by the efferent olivo-cochlear bundle may serve as the neuronal basis for the critical band phenomenon. This gating effect by neuronal inhibition has been clearly demonstrated to exist (cf. Rasmussen, 1946; Desmedt and Monaco, 1961; Fex, 1962). This suppression of neural response may occur at:

1. individual afferent fibers of inner hair cells (post-synaptic), and

2. the entire output of individual outer hair cells (pre-synaptic) (Spoendlin, 1977).

Fex (1967) proposed a network describing the afferent and efferent pathways involved with each cochlea. From this morphology, a hypothetical feedback mechanism may be proposed (see Figure 1). The ascending pathway courses from the cochlea's hair cells through the VIII nerve to the cochlear nucleus and then to the superior olives via the olivo-cochlear bundle to the cochlea. Since there is a 33:1 ratio of afferent to efferent fibers, there is a limit to which frequency information may be gated. Scharf (1970) reported twenty-four critical bands (narrow frequency regions of gated information) whose widths equal approximately 16 percent of center frequency for those bands above 500 Hz (Table 1).

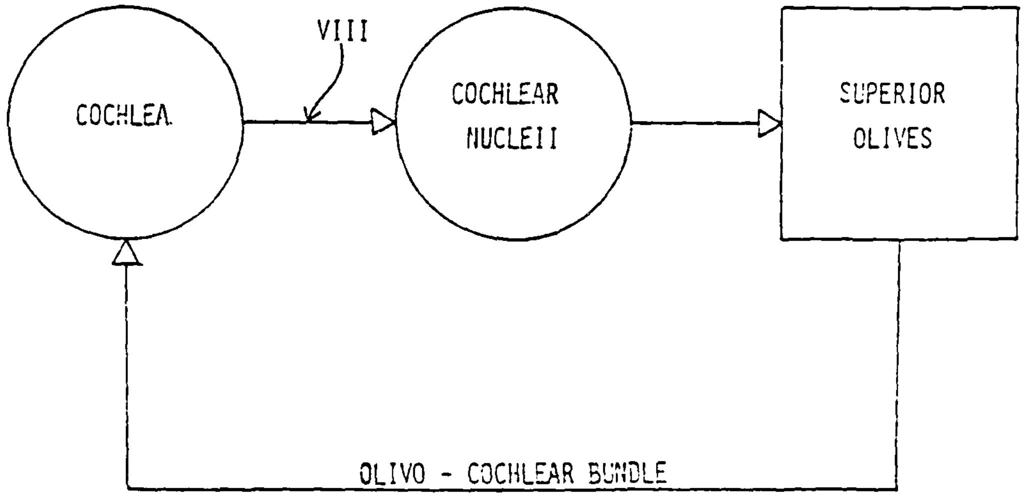

Figure 1. Hypothetical Feedback Mechanism.

It would follow that any decrement in the efficiency of this feedback mechanism would incur deficits in the functional characteristics of the critical band mechanism described above. Several studies (Fex, 1967; Dewson, 1968; Capps and Ades, 1968; Trachiotis and Elliott, 1970; Pickles and Comis, 1973) have examined the behavioral performance of animals in which the action of the olivo-cochlear bundle was blocked. Two overriding observations were made:

1. reduction in frequency discrimination ability, and

2. reduction in the ability to recognize signals in a background of noise.

Thus, it is reasonable to assume that units of the neural inhibitory system of the cochlea would be damaged whenever significant cochlear pathology is present. In such a case, one would expect to find wider-than-normal critical bands due to the loss of inhibitory units.

There is no reason to assume that the central nervous system is aware of the particular details of a peripheral pathology. Thus, the central nervous system expects to receive information from a full complement of critical bands. Hence, the widening of individual critical bands involves two factors:

1. a broadening of the integration region which results in each critical band integrating a larger area for a given signal;

2. retention of the complete number of critical bands such that the critical bands may be expected to overlap one another in the widened case.

Therefore, the energy content of frequency regions common to more than one band will be integrated more than once. This phenomenon may be expected to lead to an abnormally high perception of loudness for a given magnitude of input acoustic energy. This phenomenon is oberved psychoacoustically among sensorineural hearing impaired individuals, and is referred to as recruitment (cf. Fowler, 1928; Michael and Bienvenue, 1976).

Bonding (1979) observed indications of widened critical bands in some 50 - 67 percent of the sensorineural hearing impaired listeners which he examined. In addition, Bonding's data demonstrate that the width of the widened critical band is independent of the magnitude of threshold hearing loss amongst those sensorineurals with critical bandwidth distortion. This finding has been supported in tests by Michael and Bienvenue (1976); Bienvenue and Michael (1977); and Bennett, et al., (1978), who found evidence of widened critical bands in noise-exposed patients which was not correlated to threshold shift magnitudes. In fact, some critical bandwidth distortion may occur in the absence of threshold shift (Michael and Bienvenue, 1976).

The two functional characteristics of this mechanism suggest the symptoms that may appear in a cochlear pathology. A common finding among individuals with cochlear hearing loss is that relatively small amplitudes and remote frequencies of background noise are detrimental to speech perception. This phenomenon may very well follow directly from the reduced band-limiting capabilities of the widened critical bands (Michael and Bienvenue, 1976). In addition, it is clear that such a pathology, resulting in a widening of critical bands, will tend to reduce the number of discriminable harmonics of a signal such as speech. Listeners with this problem will be less able to discriminate speech on the basis of its harmonic content. Subjects with cochlear pathologies report speech to sound "foggy" or "blurred," with some insisting that everyone mumbles when they speak (cf. Fowler, 1928; 1937). The integrity of the listeners' critical bands, therefore, appears to represent a limiting factor in their ability to perform these everyday tasks.

The plight of this pathology lies in its lack of remedial measures. While this phenomenon has existed amongst the general populace for as long as the classic case of threshold loss, no present-day audiometric aids are available that can alleviate these symptoms. Simply amplifying the signal to try to compensate for such deficiencies only worsens the effect by presenting both the desired speech and the unwanted background noise equally loud. Rather, a means to test for early signs of bandwidth widening might prevent extensive damage and provide a pool of knowledge on which to base remedial measures. Standard laboratory procedures for measuring critical bandwidth typically involve lengthy psychophysical techniques performed on trained subjects (e.g., Fletcher, 1940; Zwicker, 1954; Greenwood, 1961; Haggard, 1974). These procedures, while extremely powerful and accurate, are not easily applied to the clinical setting where listeners are untrained and unwilling to spend the requisite time listening to sophisticated signal complexes. Development of a clinical testing procedure which determines critical bandwidth directly and rapidly should prove more desirable than the standard tests.

Chapter II

STATEMENT OF THE PROBLEM

The condition of widened critical bands should present a disruption to the source-path-receiver communication chain. Given a clearly articulated speech signal propagating through a conducive acoustic medium, one finds a receiver whose frequency "resolving power" is insufficient to enable discrimination of that speech from the entire acoustic stimulus. The degree to which the listener has lost his/her resolving power is a variable (N = factor by which the critical band is widened, see Figure 2a). The approach chosen to determine N involves adding an interruptive processor to the communication chain, whose resolving characteristics are known and can be varied at will (see Figure 2b). By observing the response of the unknown receiver while varying the known disruption of the presentation, N may be determined. If the frequency resolution of the interruptive processor is finer than the resolving power of the listener, then perception is limited simply by the listener's critical bandwidth. Widening the bandwidth of frequency resolution for the processor should have no effect upon the listener's acoustic analysis of the signal (compared to the unprocessed case) until the resolution becomes coarser than the listener's critical bandwidth (i.e., his/her frequency resolution capacity), at which point a decrement in performance should be observed. Thus, the

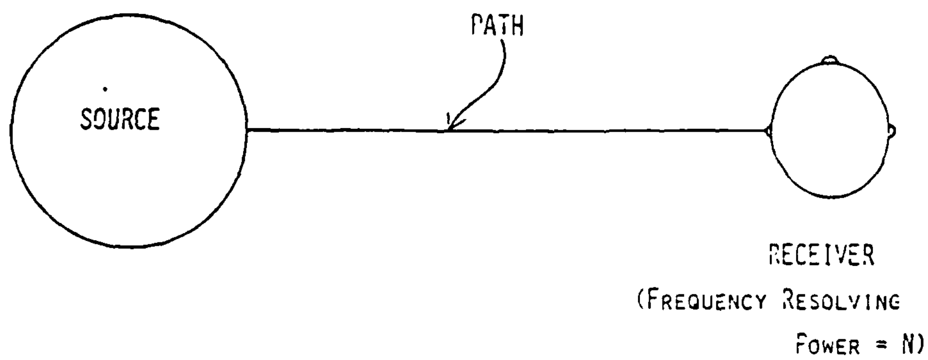

FIGURE 2A. COMMUNICATION CHAIN.

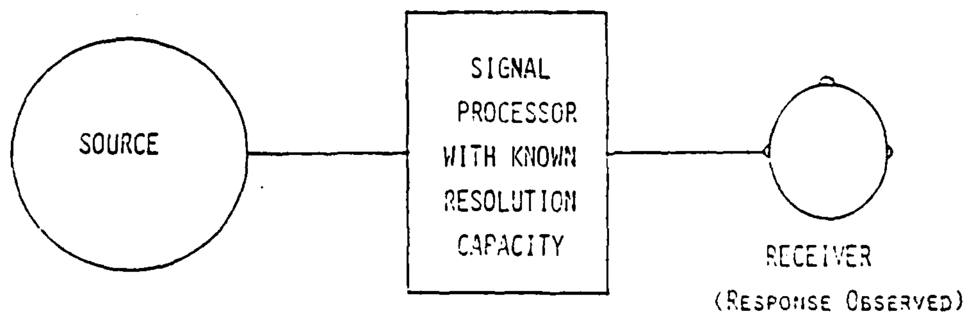

FIGURE 2B. COMMUNICATION CHAIN WITH INTERRUPTIVE PROCESSOR.

critical bandwidth of a listener in this procedure equals N-times the normal bandwidth capability, indicated by the widest resolution at which he/she exhibits non-decremental performance trends.

The purpose of this study, then, is to:

1. Generate acoustic stimuli of varying resolutions equal-to- or greater-than-normal critical bandwidth using digital signal processing; specifically,

    a. Is it feasible to produce bandwidth-limited signals such as those described in Chapter III?

    b. Is it feasible to produce a processing scheme which allows for variation of bandwidths from the normal critical band to integer multiples of the critical band?

    c. Do these varying bandwidth-limited speech signals demonstrate characteristics observed in speech processing by humans under pathologic conditions such as the phenomenon of recruitment?

2. Using these signals as stimuli for speech discrimination testing with normal listeners, specifically,

    a. Do normal listeners, presented with narrower-than- or equal-to-normal resolution limited signals (including non-bandwidth-limited signals), demonstrate statistically non-significant performance trends for these conditions? That is, do narrower-than- or equal-to-critical bandwidth limited signals obscure no more spectral content needed for speech discrimination than does the filtering action of a normal auditory system?

b. Do normal listeners, presented with wider than normal bandwidth resolution limited signals, demonstrate performance decrements comparable to those seen in sensorineural hearing impaired listeners? That is, does a widened bandwidth condition effectively model sensorineural hearing impaired speech listening?

c. Do normal hearing listeners demonstrate a monotonic trend of decreasing performance in speech intelligibility as their allowed bandwidth resolution is systematically widened? That is, does the magnitude of bandwidth widening effectively model the magnitude of impairment in speech discrimination?

3. Using these signals as stimuli for speech discrimination testing with sensorineural hearing impaired listeners, specifically,

a. Do sensorineural hearing impaired listeners, presented with resolution limited signals that are narrower than their own pathologic bandwidth resolution capability, demonstrate statistically non-significant performance trends for these conditions? That is, do bandwidth limited signals that are narrower than the sampled sensorineural mean bandwidth capacity obscure no more spectral content needed for speech discrimination than does the filtering action of their pathologic auditory systems?

b. Do sensorineural hearing impaired listeners demonstrate a monotonic trend of decreasing performance in speech intelligibility as their allowed bandwidth resolution is systematically widened beyond their own bandwidth resolution capability?

4. Measure the critical bandwidth of the normal listeners by an independent procedure, specifically,

a. Is it feasible to generate tonal complexes that systematically widen with time from a sub-critical to a supra-critical bandwidth, such as those described in Chapter IV?

b. Do normal hearing listeners demonstrate a mean bandwidth rating that is statistically equal to those found in other psychoacoustic studies?

c. Are the performance trends of this independent critical band test for normal listeners correlated to their performance trends for the speech discrimination test? That is, is their mean critical bandwidth rating, expressed as a decimal multiple of one critical band, statistically equal to the widest resolution at which statistically equal mean performances occurred in their speech discrimination test?

5. Measure the critical bandwidth of the sensorineural hearing impaired listeners by an independent procedure, specifically,

a. Are the performance trends of this independent critical band test for sensorineural hearing impaired listeners correlated to their performance trends for the speech discrimination test? That is, is their mean critical bandwidth rating, expressed as a decimal multiple of one critical band, statistically equal to the widest resolution at which statistically equal mean performances occurred in their speech discrimination test?

Chapter III

DIGITAL SIGNAL PROCESSING OF SPEECH MATERIALS

General Overview

The generation of bandwidth resolution limited speech signals involves the digital signal processing algorithm shown in Figure 3. Prerecorded stimuli (in analog form) are digitized by a computer, digitally processed, and then rerecorded onto audio tape in a processed analog form. The first step of this procedure requires the transformation of a continuous input into a series of discrete elements. The input has a continuously varying amplitude and is the analog form of the signal. The digital form of the signal contains an array of discrete values corresponding to the input amplitude. The device used to transform the signal from a continuous to a digital mode is called an analog-to-digital (or A/D) converter. Within this converter is a timing pulse which beats at a fixed rate known as the sampling rate. As the analog signal is fed into the A/D converter via a conventional playback machine, the amplitude of the signal is detected at every occurrence of the timing pulse. The digital portion of a computer may receive and store these discrete amplitudes of the wave in the form of a one-dimensional array of voltages. For example, a digitized sine wave might have an array "A" equal to (0, 2, 4, 6, 4, 2, 0, -2, -4, -6, -4,...). The sampling rate commonly used for audio signals is in excess of 20,000 samples per second.

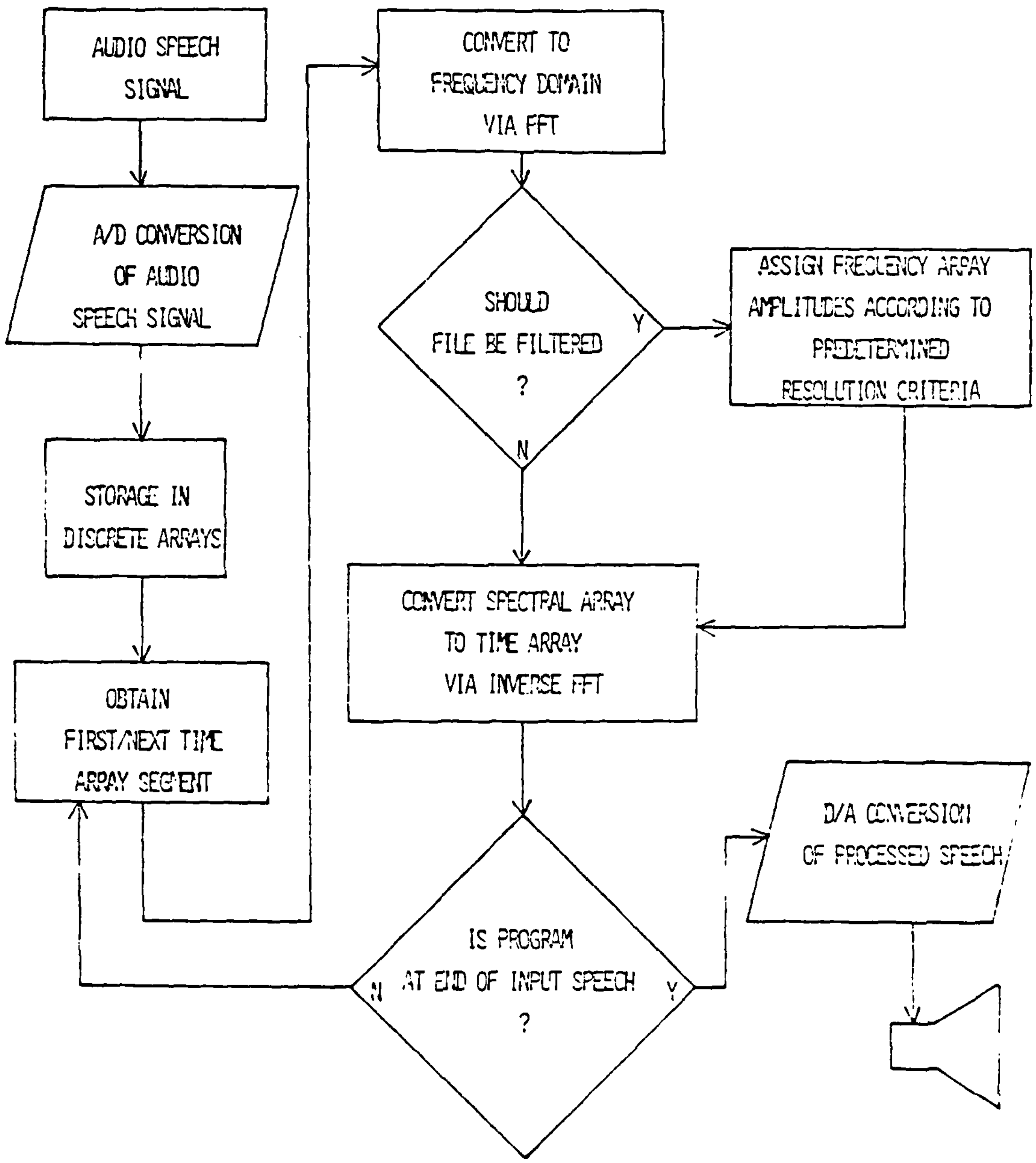

Figure 3. Processing Algorithm.

Thus, the example given above would be a typical array representing a pure tone at or above 2000 Hz (depending on the precise sampling rate).

Once the word list is digitized and stored as a time domain array within a digital computer, the processing scheme may be initiated by the software program. The data are then converted in small time increments into the frequency domain by what is known as a Fast Fourier Transform.

The Fast Fourier Transform

A convenient and precise method for analyzing audio signals involves delineation by sums of sinusoids or complex exponentials. Commonly called Fourier representations, they provide an inherently superior tool to signal processing for two fundamental reasons. First, a linear system's response may be easily determined from a superposition of sinusoids or complex exponentials. Secondly, the Fourier representation often reveals properties of a signal that would otherwise be less evident (Rabiner and Schafer, 1978).

Early models for speech production of steady-state vowels or fricatives, for example all involved a linear system excited by either a periodic or random source. Consequently, Fourier analysis was utilized traditionally in the evaluation of such spectra. More recently, however, speech has been viewed as a much more dynamically complicated waveform (cf. Ladefoged, 1962; Curtis, 1968; Minifie, et al, 1973). The combined transient, random and periodic nature of a speech signal induces marked changes in amplitude with time, violating the steady-state requirements of a standard Fourier representation. Instead, a short-time analysis principle applied to the Fourier method has been found to be a valid approach to speech processing (Rabiner and Schafer, 1978). These authors found that a steady-state assumption for the spectral properties of speech is valid for time intervals on the order of 10-30 msec. In a review of this time-varying technique, the application of its principles to fast computation algorithms for discrete Fourier analysis (FFT algorithms) will be demonstrated.

Classical Fourier analysis of spectra has two basic approaches. For purely periodic waveforms, one determines its Fourier Series (see Equation 1):

$$X(t) = a_o + \sum_{m=1}^{\infty} \{a_m \cos(m\omega_o t) + b_m \sin(m\omega_o t)\} , \qquad (1)$$

where: t = time,

T = the period of $X(t)$,

$\omega_o = 2\pi/T$,

$$a_o = 1/T \int_0^T X(t)dt ,$$

$$a_m = 2/T \int_0^T X(t)\cos(m\omega_o t)dt ,$$

and

$$b_m = 2/T \int_0^T X(t)\sin(m\omega_o t)dt .$$

By contrast, pulse-like waveforms are analyzed by evaluation of its Fourier Transform (see Equation 2).

$$x(f) = \int_{-\infty}^{\infty} X(f)\exp(-j2\pi ft)dt \qquad (2a)$$

or

$$X(t) = \int_{-\infty}^{\infty} x(t)\exp(-j2\pi ft)df \ . \qquad (2b)$$

Underlying the Fourier Series method is the following notion: the elemental periodic waveform is a sinusoid of the form:

$$x(t) = A\cos(2\ ft - \theta)\ .$$

Further, all periodic waveforms are consequently comprised of some unique summation of sinusoids. Each of the summation terms exists at the discrete frequencies given by $\{m\omega_o\}$. Each term is also harmonically related to the fundamental, 1/T by the index m. The two parameters that define that set are the amplitude spectrum and the phase spectrum. Whenever these two parameters are evaluated, whether electrically, mechanically, or mathematically, the process said to be occurring is called spectral analysis.

The Fourier Transform, on the other hand, defines x(f) as a continuous function of frequency. There is no index, m, as in the Fourier Series, which would have indicated a dependence upon discrete frequencies. Aperiodic, or pulse-like waveforms must have their complete time history integrated to determine the corresponding frequency composition.

While analyzing speech, however, one finds a mixture of both periodic and aperiodic waveforms (cf. Minifie et al., 1973); neither method alone is complete. Rather, a Discrete Fourier Transform (DFT) capitalizes on the discrete nature of the waveform's amplitude to enable provision of spectral information regardless of periodicity (see Equation 3):

$$X\ (k\ f) = (1/M\Delta t) \sum_{m=0}^{M-1} x(m\Delta t)\exp(-j2km/M)\ . \qquad (3)$$

where: t = time,

T = time interval of sampling,

Δt = time sampling spacing,

Δf = frequency sampling interval = $1/T = 1/Mt$,

M = number of samples in T,

m = time index (0, 1, 2, ..., M-1) ,

and

k = frequency index (0, 1, 2, ..., M-1) .

If the source function $x(m\Delta t)$ repeats itself with time, the evaluation occurs as it would in a Fourier Series computation. In the case of a transient function, the array of distinct amplitude values capacitates a direct summation of the complex Fourier Transform. It should be noted that computation algorithms involving a series summation are much more efficiently realized by a computer than are formal evaluations of integrals. Thus, the DFT, which serves as the basic algorithm of a Fast Fourier Transform (FFT), efficiently performs spectral analysis of speech signals, given adherence to certain necessary criteria, described below.

The main requirement for using the DFT is that the digitized speech waveform must satisfy the Nyquist sampling criterion; the sampling rate should be at least twice as great as the highest frequency in the waveform sampled (Rabiner and Schafer, 1978). Sampling at twice the highest frequency contained in the input provides greater than two samples for each fundamental waveform; this insures proper coding of the signal's frequencies. Violation of this rule leads to the undesirable phenomenon of aliasing, in which high frequency amplitudes are confused as low frequency information (see Figure 4). Foldover is a term which describes the magnitude of frequency displacement error induced by the aliasing phenomenon. That is, half the sampling frequency serves as a pivot frequency for aliasing in that the low frequency alias occurs at a frequency as far below the pivot frequency as the high frequency component is above the pivot frequency. For example, if a sampling rate of 20 kHz is used for speech (yielding a pivot frequency of 10 kHz), any high frequency component at 15kHz is "folded down" to become a 5000 Hz low frequency alias, yielding an inaccurate spectrum (see Figure 5).

Another important consideration in the computation of the DFT within a software program is the time necessary to complete it. Cooley and Tukey (1965) found a significant reduction in the number of complex additions and multiplications needed for this transform whenever the number of samples chosen for each computation equaled a power of 2 (i.e.,

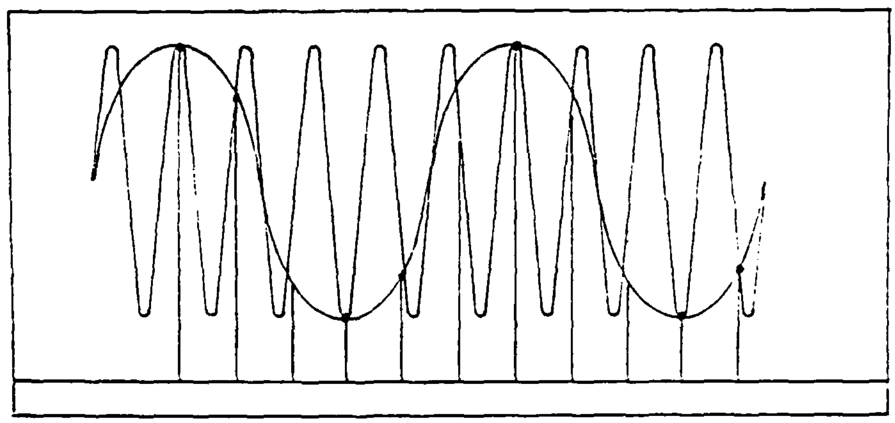

Figure 4. Aliasing Phenomenon.

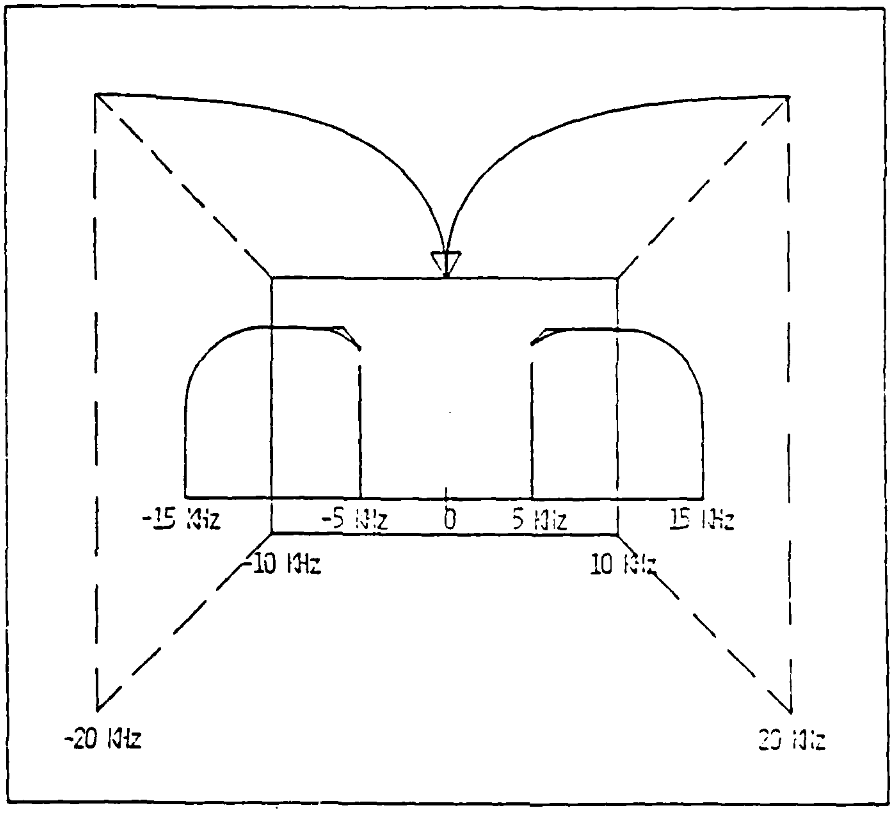

Figure 5. Foldback Phenomenon.

when $M = 2^n$). In addition, the two indices involved in the compilation always have a value of 0 or 1 for $M = 2^n$, a feature exploited by certain FFT subroutines to gain an additional time advantage (Singleton, 1969).

An elementary property of the Discrete Fourier Transform is that it is a linear operation (Cooley et al., 1969). Stated mathematically, $X_d(k\Delta f) \leftrightarrow x(m\Delta t)$, it indicates the validity in performing an inverse DFT. Since this research manipulates spectra while in the frequency domain, a viable method to gain access to <u>and</u> from that realm would be a necessary and sufficient requirement. Thus, the linearity of the DFT provides the symmetrical tool upon which such processing as digital filtering depends.

The FFT is the functional software realization of the Discrete Fourier Transform. Used as a subroutine, it makes available (in a forward transform) arrays corresponding to the real and imaginary components of a spectrum's amplitudes. Conversion of these values to polar form yields one magnitude and one phase value for each frequency array element. There is a fixed frequency interval between the source values for each array element; for example, array element number one might correspond to the amplitude of that instantaneous signal at 70 Hz, while array element number two might correspond to the instantaneous 140 Hz amplitude, etc. The number of frequency elements depends on the sampling rate used to initially digitize the waveform, and the number of samples (time segment size) used in the FFT process. Sampling rates are generally in excess of 20,000 samples/second in order to satisfy the Nyquist criterion for the primary speech audio range (i.e., below 10,000 Hz), while time segments on the order of 10-30 msec are taken sequentially to approximate the steady-state condition described earlier (cf. Rabiner and Schafer, 1978).

Note that this information is stored independent of specific frequency data. The affiliation of a voltage value with a particular frequency is an arbitrary component of the output process and this stored array of voltages is independent of frequency information prior to output processing. Thus, the term "filtering" takes on a new meaning in the digital mode. Instead of running the signal through a relatively coarse analog filter, each instantaneous spectral array may be modified by simply specifying the energy content between predetermined frequency limits (see Figure 6). The slopes on digital "filters" are nearly infinite and permit the generation of tightly tuned, nonoverlapping band-pass filtering assignments like those found in the normal auditory periphery (cf. Scharf, 1970).

The processed spectra were made using the frequency limits recommended by Scharf (1970) and reported above in Table 1. The discrete frequency amplitudes that fall within each bandwidth are averaged; each of the discrete amplitudes of that band are then set equal to this r.m.s. value, limiting the resolution allowed to the preselected bandwidth for that time segment (see Figure 6). The bandwidths shown in

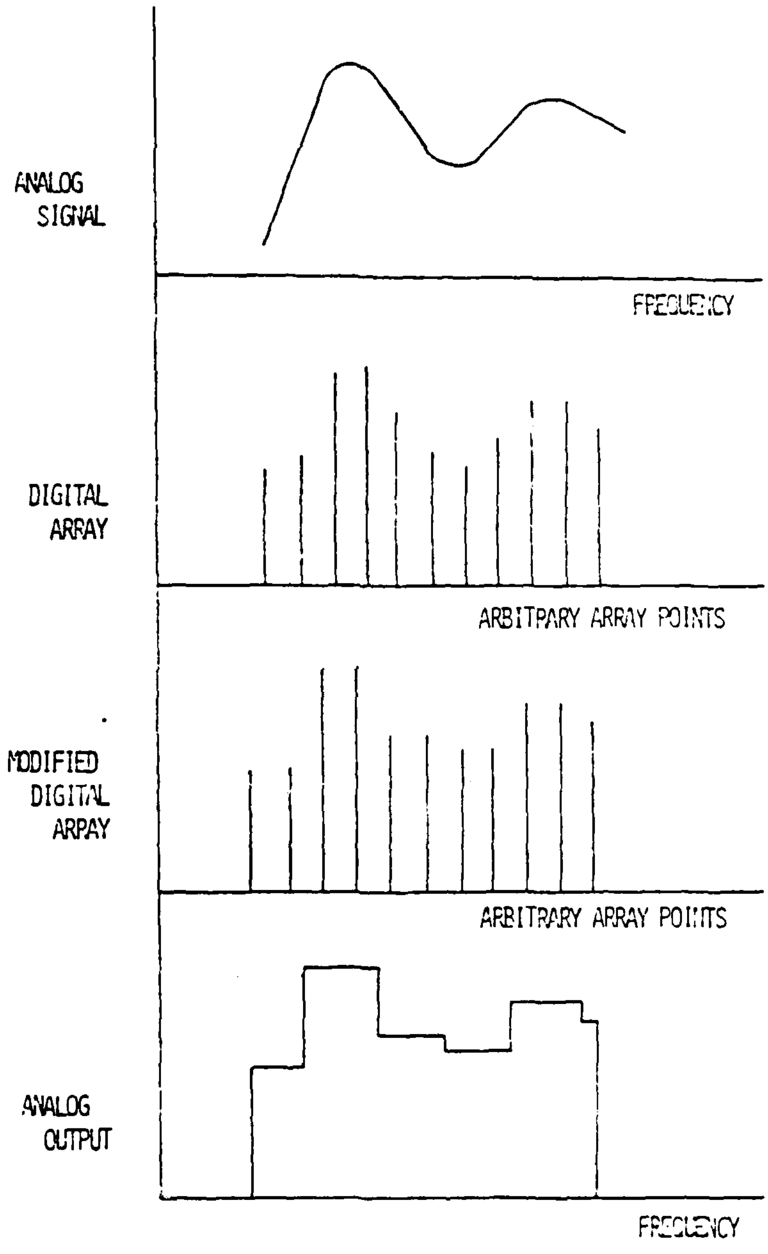

FIGURE 6. EFFECT OF DIGITAL FILTERING.

Table 1 give the limits for a normal critical bandwidth (i.e., CB = 1X). Coarser filtering schemes are realized by multiplying the bandwidths by a chosen integer value (retaining the original center frequency), and averaging the amplitudes contained within these widened limits.

Once the frequency assignments have been made, the new array of spectral magnitudes are converted to rectangular form and placed in a call to an inverse FFT.

The processed speech segment, now back in the time domain, is stored in a new array to await output. The loop (involving a conversion to the frequency domain, the implementation of frequency assignments and subsequent call to inverse FFT) continues until all time segments have been processed.

The Smoothing Process

This procedure of taking the speech signal "a slice at a time" for processing takes advantage of the discrete nature of the stored signal. Simply recompiling the string of processed segments assumes that the envelope of each time slice does not differ drastically from what it was before processing. This has not been found to be true in practice, however. In fact, substantial noise appearing in the output of such processing may result directly from this practice. Consider, for example, the demarcation point "A" in Figure 7a, indicating where one time segment ends (array point 19) and another begins (array point 20). After processing (see Figure 7b), the relative amplitudes across that juncture are

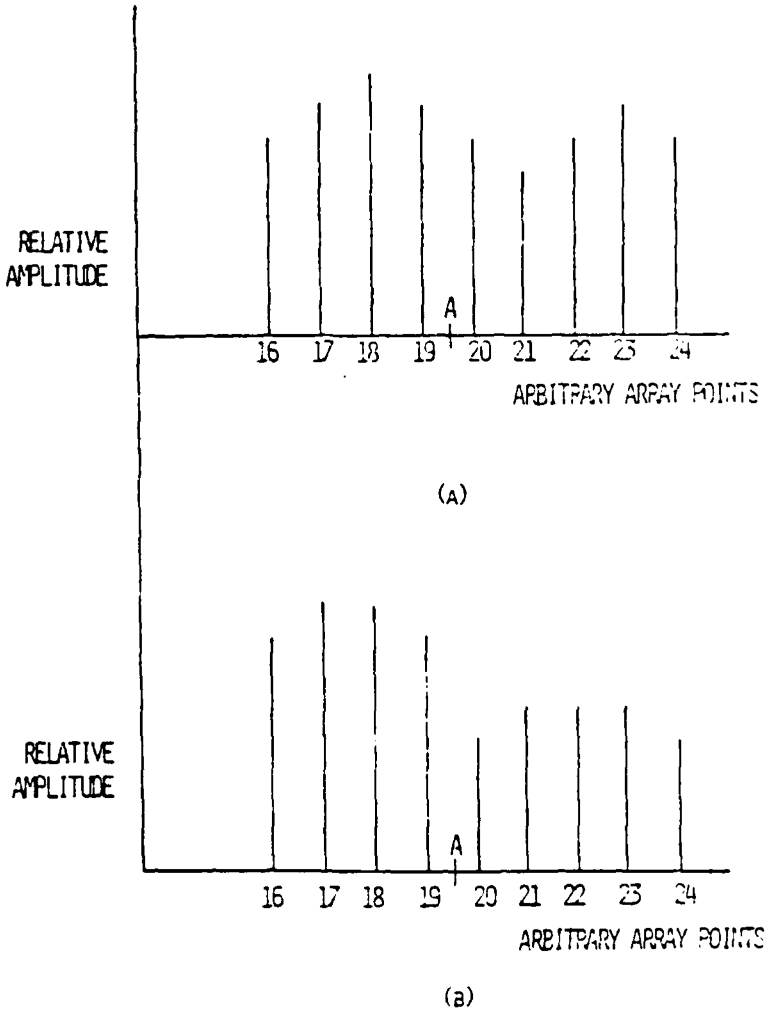

Figure 7. Example of Envelope Discontinuity across Time Segment Boundary

significantly disparate from one another due to the composite spectral changes made within each segment. Analog reproduction devices (especially earphones and loudspeakers) are unable to accurately transduce such a jump in amplitude. The resultant acoustic output at such a point is a transient "pop". If, for example, the time segments are each 15 msec long, then one transient would occur every 15 msec. This translates into a 67 Hz "buzz" signal which modulates the entire acoustic output, distorting its spectral content.

To rectify this inherent situation, a software procedure was composed which will henceforth be referred to as "smoothing." The technique basically involves an isolation procedure to prevent significant envelope changes from occurring across each processed time segment boundary. The first call to FFT (henceforth known as a "pass") sends a specific number of time domain amplitude values to be converted into the frequency domain. Upon assignment of the specified spectral shape, the data are returned to the time domain via an inverse FFT call; this pass is indentical to the general procedure described earlier. The second pass involves the same amount of array points in the FFT call as in the first pass, however, now the first 10 percent of the points are the same array points as the last 10 percent of the first pass' call to FFT. For example, if the first pass sends array points 1-256 to FFT, pass number 2 would send array points 230-486 to FFT. Further, pass number 3 would send array points 460-716, and so forth. This repetition between the values at the start and finish of each array isolates the boundary of each segment from large envelope discontinuities. After the last pass has been completed, the entire string of processed segments are rewritten to an output array by taking only the first 90 percent of each segment (except for the final pass, taken in its entirety). This accomplishes two goals: the repetition of small time sectors is edited, while a more nearly continuous envelope change across the boundary is approximated. The signal's amplitude vs time history is still discrete; however, these amplitudes now vary across the segment boundaries with inter-segment smoothness. Hence, the smoothing procedure adjusts the precise features of processing-induced transients in the signal's envelope on a software level, such that auditorily perceptual "pops" are eliminated from the output while retaining the data in digital form.

Note that these manipulations (filtering assignments and smoothing procedures) all occur outside of the signal's real time. This characteristic offers several unique advantages. First, a high degree of precision is achieved during processing of the signal. Digital editing and precise spectral shaping are examples where this feature excels. Secondly, the number of different modifications greatly increases when the signal is available as discrete quanta outside of real time. Since the entire duration of the signal is accessible as a quantified whole, all dimensions of the input may be simultaneously manipulated.

Finally, iterative schema may be conducted utilizing the speed of the computer's hardware to analyze different combinations of precise modifications. In many cases the experimenter does not have foreknowledge of the exact combinations needed to attain a specific output. A guessing procedure in real time is inherently limited by the need to completely process a signal each time a solution is tried. In the digital mode, however, the desired output is returned in one step since the iterations occur within the execution of the software program. Thus, the non-real-time nature of digital signal processing offers greater opportunity for signal modification than conventional analog filters and modulators.

Chapter IV

THE MINIMUM DISCRIMINABLE BANDWIDTH (MDB) TEST

General Overview

Independent measurement of the subject's critical bandwidths is achieved through the use of a modified method of limits technique. A tonal complex with an initially sub-critical bandwidth is presented to a listener. The test signal then discretely widens in bandwidth with time; each bandwidth has a duration of approximately 800 msec. The subjects are instructed to indicate the moment they first perceive a change in the sound. The minimum bandwidth at which a listener can discriminate a difference in the tonal complex is taken to be that subject's critical bandwidth (cf. Scharf, 1970). Thus, the procedure is called the Minimum Discriminable Bandwidth test (hereafter referred to as the MDB test).

Test Signals

The time varying tonal complexes for the MDB test are generated using the digital signal processing algorithm shown in Figure 8. Each of the signal bandwidths is generated as a discrete time segment, similar to the manner in which the speech signals were processed in the preceding chapter. However, instead of processing successive portions of an input signal, this procedure directly addresses the desired frequency content to an array already in a polar-coordinate frequency domain. Keeping in mind the stipulated

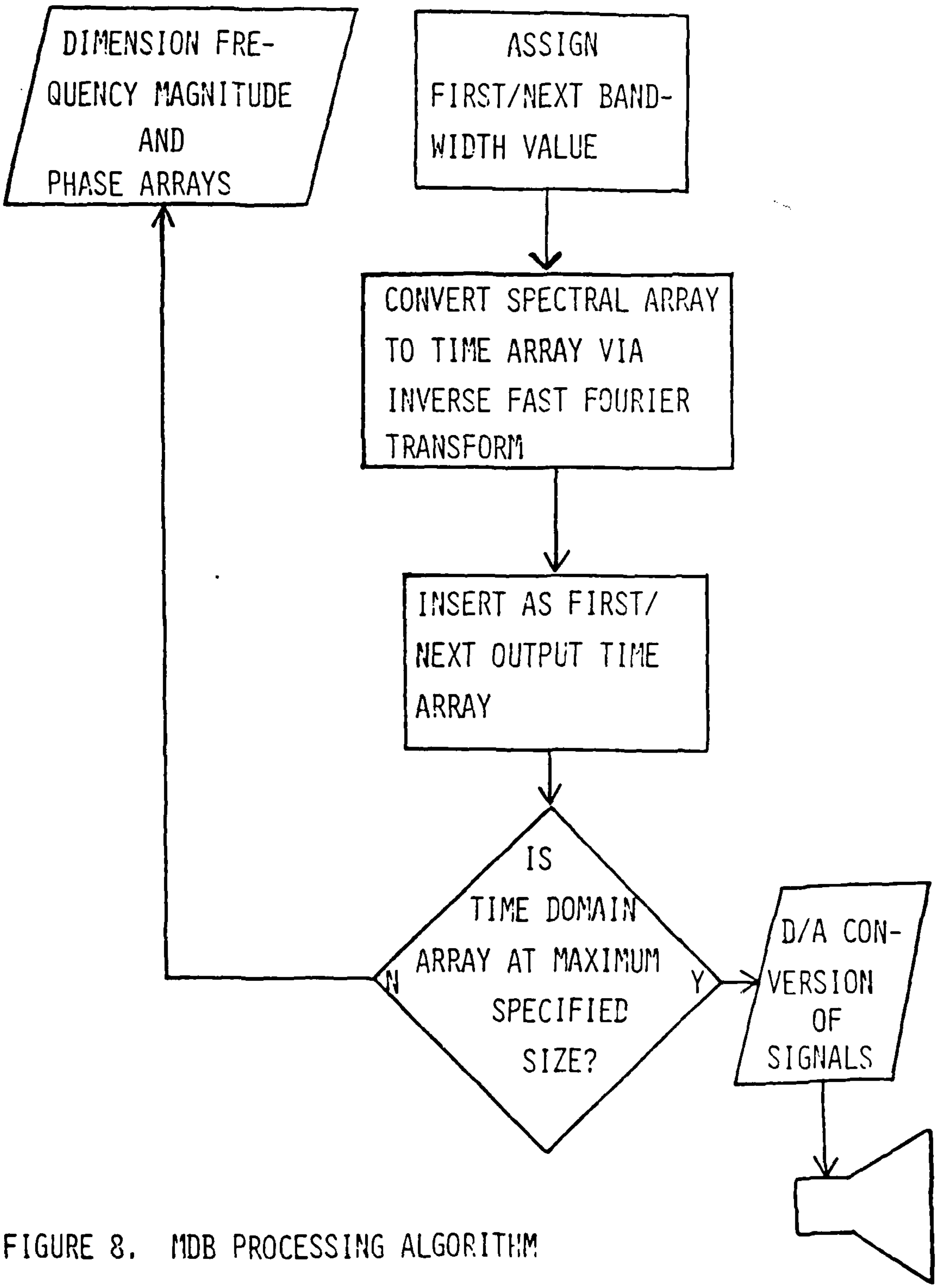

FIGURE 8. MDB PROCESSING ALGORITHM

time segment size of each signal bandwidth and the ultimate D/A sampling rate that will be used, the fixed frequency interval between each array element may be determined. Since the time segment size is inversely proportional to the frequency interval, the relatively large segment size (800 msec vs 15 msec for speech processing) results in a high frequency resolution (1.22 Hz/array point).

The amount of bandwidth increase for each time segment is chosen such that a geometric widening occurs. The magnitude values of the frequency array points utilized are kept equal to each other. In addition, these values are uniformly reduced for each successive bandwidth increase in such manner that the overall signal maintains a constant energy output.

Once the magnitude and phase relationships for a signal bandwidth are specified, these arrays are converted to rectangular form and subsequently an inverse FFT is performed. This process of successively specifying geometrically widened bandwidth values and transforming each to the time domain yields an output string of discrete time segments, each with their own bandwidth value. A demonstration of the bandwidth-time relationship for these signals is given in Figure 9. The discrete time domain output array is then recorded onto audio tape via a D/A converter.

The inherent processing limitations of this procedure are much the same as those described in the speech processing chapter, with one exception. The Nyquist criterion

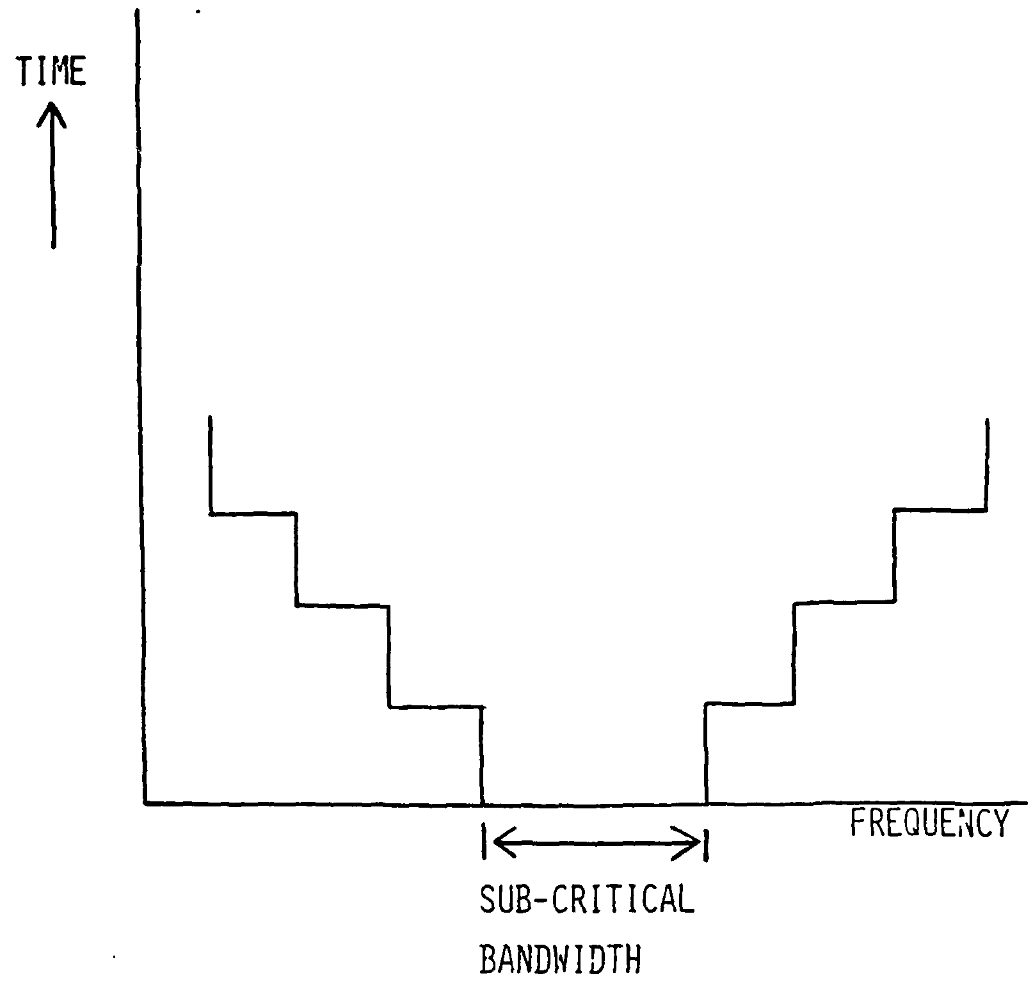

FIGURE 9. BANDWIDTH-TIME RELATIONSHIP FOR MDB SIGNALS

and time segment rules still apply, but the smoothing process is not needed for two reasons. First, each time segment is over 800 msec long, and therefore any transient envelope changes occur at a rate of only 1.22 Hz. Secondly, the amplitude of each successive bandwidth signal has a constant energy output, and abrupt envelope changes are thereby minimized. Thus, the larger time segment sizes and the constant energy characteristic of the signal envelope eliminate the need for the smoothing process.

The features of the MDB test have been described. This processing algorithm is capable of producing precise bandwidth controlled tonal complexes for use in psychoacoustic testing.

Chapter V

PROCEDURES

Subjects

Normal Hearing Group. Forty-eight subjects, thirty-one female and seventeen male, participated in the speech discrimination MDB tests. Ages ranged from nineteen to thirty-one years and were selected from the student population at the University. Subjects were screened for normal hearing, and those with thresholds greater than 20 dB at any frequency from 250 to 8000 Hz in octave intervals in both ears were eliminated. The right ear was the test ear for these subjects, unless the left ear showed the only normal sensitivity.

Sensorineural Hearing Impaired Group. Twenty-four subjects, nine female and fifteen male, participated in the speech discrimination and MDB tests. Ages ranged from thirty-one to seventy-five years and were selected by reviewing the records of The Pennsylvania State University's Speech and Hearing Clinic. Subjects were screened for their speech reception threshold (SRT), and only those with an SRT between 24 dB HL and 60 dB HL were retained. The upper limit of 60 dB HL was chosen to limit the severity of hearing impairment among these sampled listeners.

All subjects were paid an hourly wage, the amount determined by the current going rate for experimental subjects at the University.

Equipment

The processed signals described in the two previous chapters were generated using a hybrid computer at The Pennsylvania State University, University Park, Pennsylvania. The system is comprised of an EAI (Electronic Associates Incorporated) Model 680 analog computer interfaced with a DEC (Digital Equipment Corporation) digital computer, Model PDP-10. The A/D and D/A conversions were both performed at a rate of 20,000 samples/second. The audio output was recorded onto Scotch Brand #208-1/4-1200 Low Print magnetic tape via a Crown Model BP824 one-quarter-inch/half-track tape recorder at seven-and-one-half inches per second (i.p.s.).

The discrimination tasks were performed using the apparatus diagrammed in Figure 10, including an Ampex AG-440B one-quarter inch/half track tape recorder, a Maico Model MA-18 audiometer calibrated to ANSI 1969 standards, and a TDH-39 earphone fitted with an MX-41/AR cushion. The tests were performed in a Suttle Corporation Model B1 quiet room.

Taped Stimulus Materials

A clinical audiometric word list was required as the input audio material for the speech processing algorithm. Northwestern's NU#6 word list was found to be most desirable, since it includes CCNC (consonant-consonant-nucleus-consonant) sounds as opposed to only CCVC (consonant-consonant-vowel-consonant) sounds (cf. Tillman and Carhart, 1966). In other words, it contains vowel sound combina-

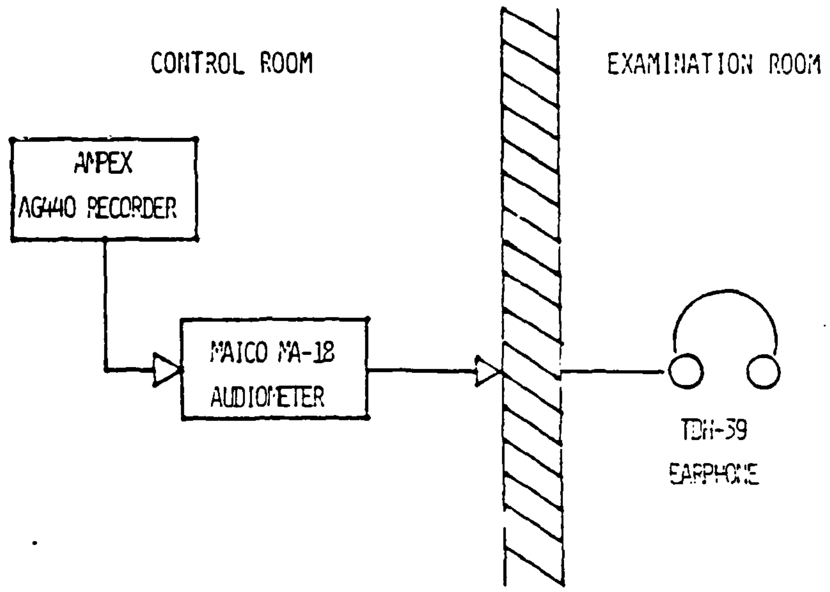

Figure 10. Experimental Apparatus.

tions, i.e., nucleii such as the /ə/ with /i/ in the word "bail", and the /a/ with /i/ in the word "bite." These nuclei occur frequently in spoken English and a word list which includes these is especially representative of the variety of sounds naturally occurring in the language.

The tapes generated have nine frequency resolutions above 500 Hz: an unprocessed list (UP); a bandwidth equal to one-half the resolution of the normal critical band (HX); a bandwidth equal to the resolution of the normal critical bandwidth (1X); one point three times the normal critical band (1.3X); one point seven times the normal critical band (1.7X); two times normal (2X); three times normal (3X); five times normal (5X); seven times normal (7X); and nine times normal (9X). The effect of the processing may be seen visually in Figures 11, 12, 13, and 14, including a comparison plot of the spectrum vs time for an unprocessed item.

Method

All subjects read and signed an informed consent document, which contained an explanation of the purpose and procedure of the study as well as an assurance of confidentiality of the data with regard to their identity. It was explained that the test basically involved listening to: 1) a standard clinical speech discrimination word list which had been modified by a novel computer manipulation technique, and 2) a set of computer-generated tones.

Speech Discrimination Test. After the screening procedure (performed in the quiet room), the subjects were

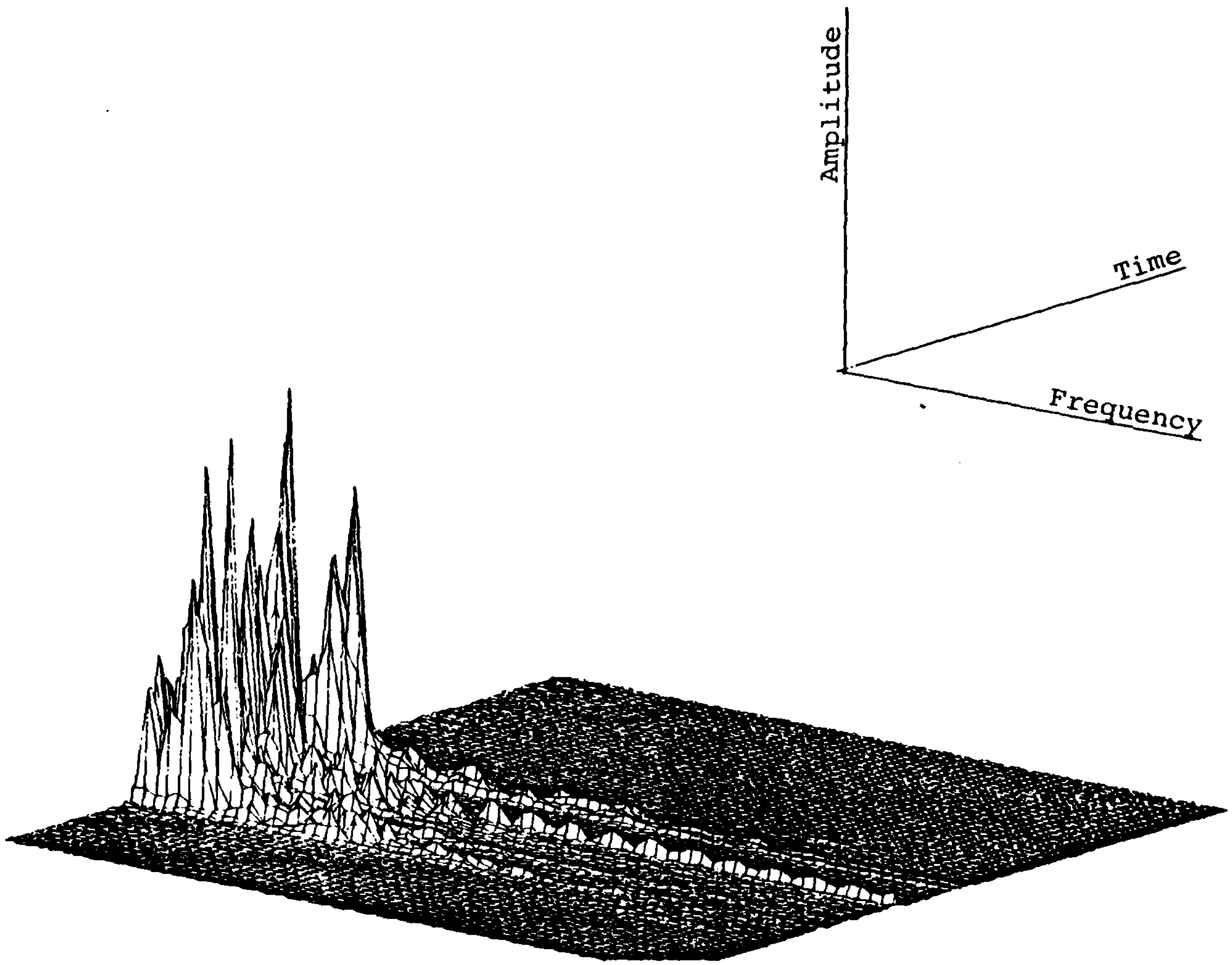

Figure 11. Three-dimensional Plot of Unprocessed Phrase.
17-Apr-80 20:06

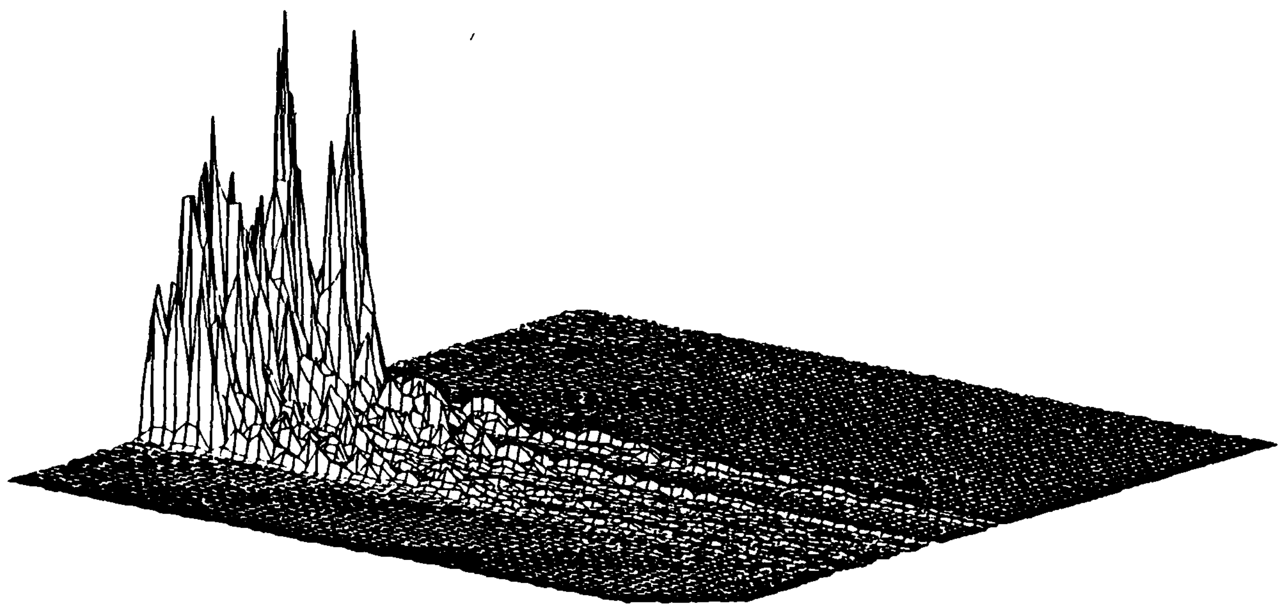

Figure 12. Three-dimensional Plot of 1X Condition.
17-Apr-80 20:31

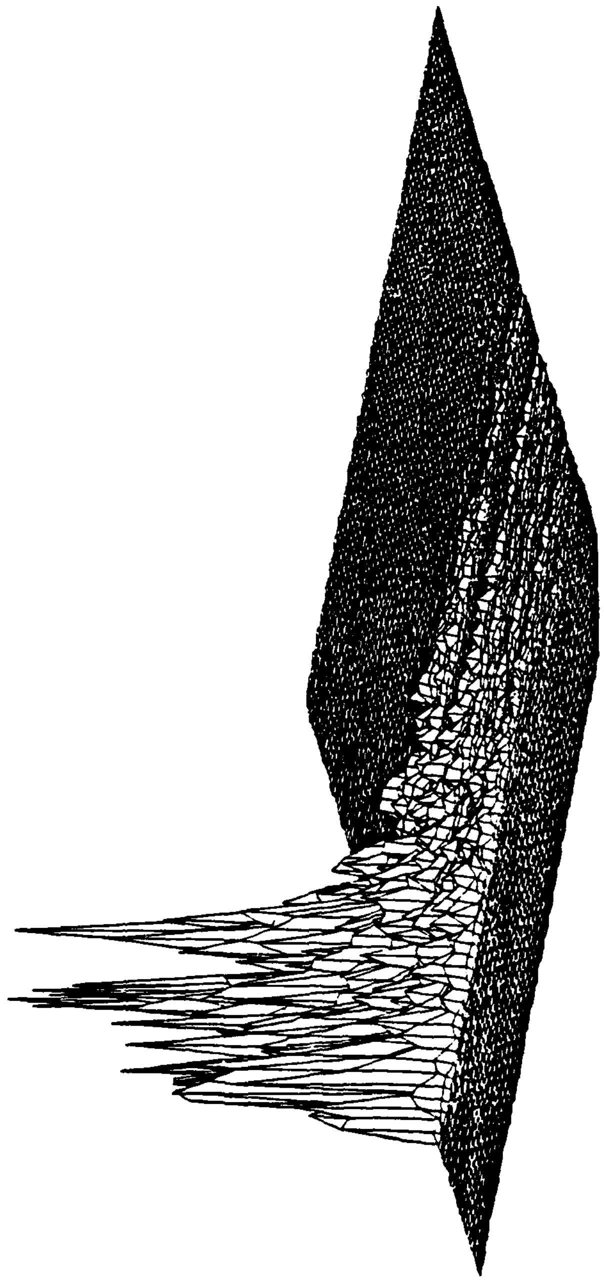

Figure 13. Three-dimensional Plot of 3X Condition.
17-Apr-80 20:51

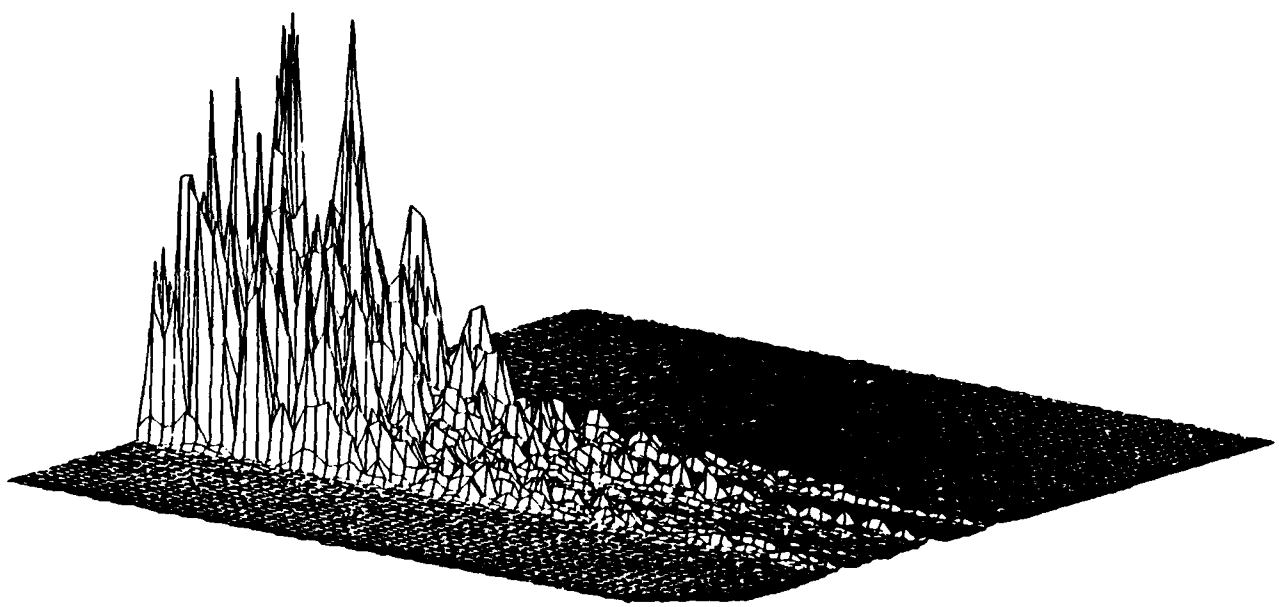

Figure 14. Three-dimensional Plot of 7X Condition.
17-Apr-80 21 36

presented with the ten fifty-word lists in the following resolution order: 9X, 7X, 5X, 3X, 2X, 1.7X, 1.3X, 1X, HX, UP. This sequence was chosen in order to minimize learning effects. The signal reached the earphone at a level of 50 dB HL and a signal-to-noise ratio of +10 dB. Pink noise was utilized as the masking source. Subjects were provided with three separate answer sheets and asked to write down the word they felt was said, guessing when necessary. Two points were scored for each correct identification, zero for an incorrect or blank answer. Each subject thus had ten percentage scores, one for each list.

MDB Test. Subjects were presented with four sets of MDB signals, with center frequencies located at 700 Hz, 1000 Hz, 1600 Hz and 2150 Hz. Ten repetitions of each center frequency were performed while the signals at the earphone reached a level of 50 dB HL. Subjects were asked to listen to each test signal episode, and indicate the moment they perceived a change in the stimulus. The amount of time that transpired from the onset of the tones to the point the subjects indicated a perceptual change was recorded for each of the trials. Each subject thus had forty duration values, ten for each center frequency.

Chapter VI

RESULTS

By comparing the plots of selected processed phrases (see Figures 11, 12, 13, 14), the effect of the processing may be seen visually. The reader should keep in mind that in all pathologic conditions (only the 3X and 7X conditions are plotted), the bands overlap each other since those in the 1X case are edge to edge. As elucidated in Chapter I, this overlap of bands introduces more area to each spectral time segment (compare Figure 11 with Figure 14), since those discrete energy regions are integrated more than once. The computer interprets this effect as adding more amplitude area to a three-dimensional array. Similarly, the auditory system perceives this effect as a loudness increase. In addition, the peaks of these processed phrases are more noticeably rounded in the 3X and 7X filtering conditions (Figures 13 and 14) than for the unprocessed case (Figure 11), indicating the reduced resolution for the widened critical band conditions. Thus, three-dimensional plots of various bandwidth-limited speech arrays exemplify characteristics indicative of the signal processing performed on the speech lists.

Pure-Tone Audiometric Screening Test

The normal group's mean threshold value for the 500, 1000, 2000, and 4000 Hz test frequencies averaged 4.1 dB HL; The sensorineural hearing impaired group's threshold value for these frequencies averaged 48.0 dB HL. The pure-tone audiogram means for both groups are reported along with the corresponding range of scores in Appendix A.

The MDB Test

The group means and variances for the MDB test are given in Table 2. These means represent the average length of time (in seconds) that transpired from the onset of the tonal complex until the first perceptual change in the signal was reported by the listener. Since the entire test signal widens in bandwidth from sub-critical to supra-critical at a fixed discrete rate, it is possible to convert these mean durations into decimal multiples of one critical bandwidth (cf. Scharf, 1970) (see Table 2). Thus, the MDB test indicates the narrowest bandwidth (re: one critical band) at which a listener first reports a perceptual change.

The MDB test indicated different critical band measures for the two groups; the normals averaged 0.90 times one critical band (0.90X), while the sensorineural hearing impaired group attained a larger bandwidth of 1.82 times one normal critical band (1.82X). A 't' test for independent groups with unequal sample sizes indicated that the mean value for the sensorineural hearing impaired group was significantly larger than the normal group mean for alpha equal to .05 (see Table 3).

Table 2

MDB Test Results

	NORMALS	SENSORI-NEURALS
Mean Response Time (seconds)	4.5	8.0
Standard Deviation	0.9	1.4
Critical Bandwidth Rating (re: One Critical Band)	0.90X	1.82X

Table 3

Behrens-Fisher t-Test
for Critical Bandwidth Scores

Mean1 = 4.5 (sec)	Mean2 = 8.0
S.D.(1) = 0.9	S.D.(2) = 1.4
n(1) = 48	n(2) = 24
c = 0.17145	df' = 32.8
t' = 11.148*	t(critical) = 2.042

* significant at .05 level

The Discrimination Test

The group means for the discrimination test are given in Table 4. The cell means represent the average discrimination score attained during each bandwidth-controlled presentation. The inordinately high mean speech discrimination score for the 1.3X condition for both groups prompted a re-examination of the recorded signal levels on the entire test tape. It was found that the keywords of the 50 phrases processed under the 1.3X condition were recorded at an average level of +3.4 dB above the 1000 Hz reference tone, while the other lists averaged only 1.1 dB above the reference tone. Therefore, the 1.3X condition list of the speech discrimination test was conducted under a less rigorous signal-to-noise ratio than were the other processed lists. It appears that the inadvertently high stimulus level for this particular test condition is the reason for its disproportionately high speech discrimination scores for both groups. Note that these scores were high only in relation to the other test condition scores. Because of this recording error, the speech discrimination scores for this condition have been eliminated from all subsequent data analyses.

The plotted means of the two groups are presented in Figure 15. A regression line was computed for the 2X through 7X portion of both sets of points. The bandwidth condition vs the normal group's discrimination scores exhibited a correlation of $r = -.75$. The regression line for

Table 4

Mean Discrimination Scores (percent) for
Tested Bandwidth Resolutions

	UP	HX	1X	1.3X	1.7X	2X	3X	5X	7X	9X
NORMALS	78.1	75.7	78.9	83.6	72.8	70.2	59.8	49.8	43.2	21.2
SENSORI-NEURALS	52.2	52.9	53.5	60.9	46.8	41.8	35.2	25.7	24.0	7.5

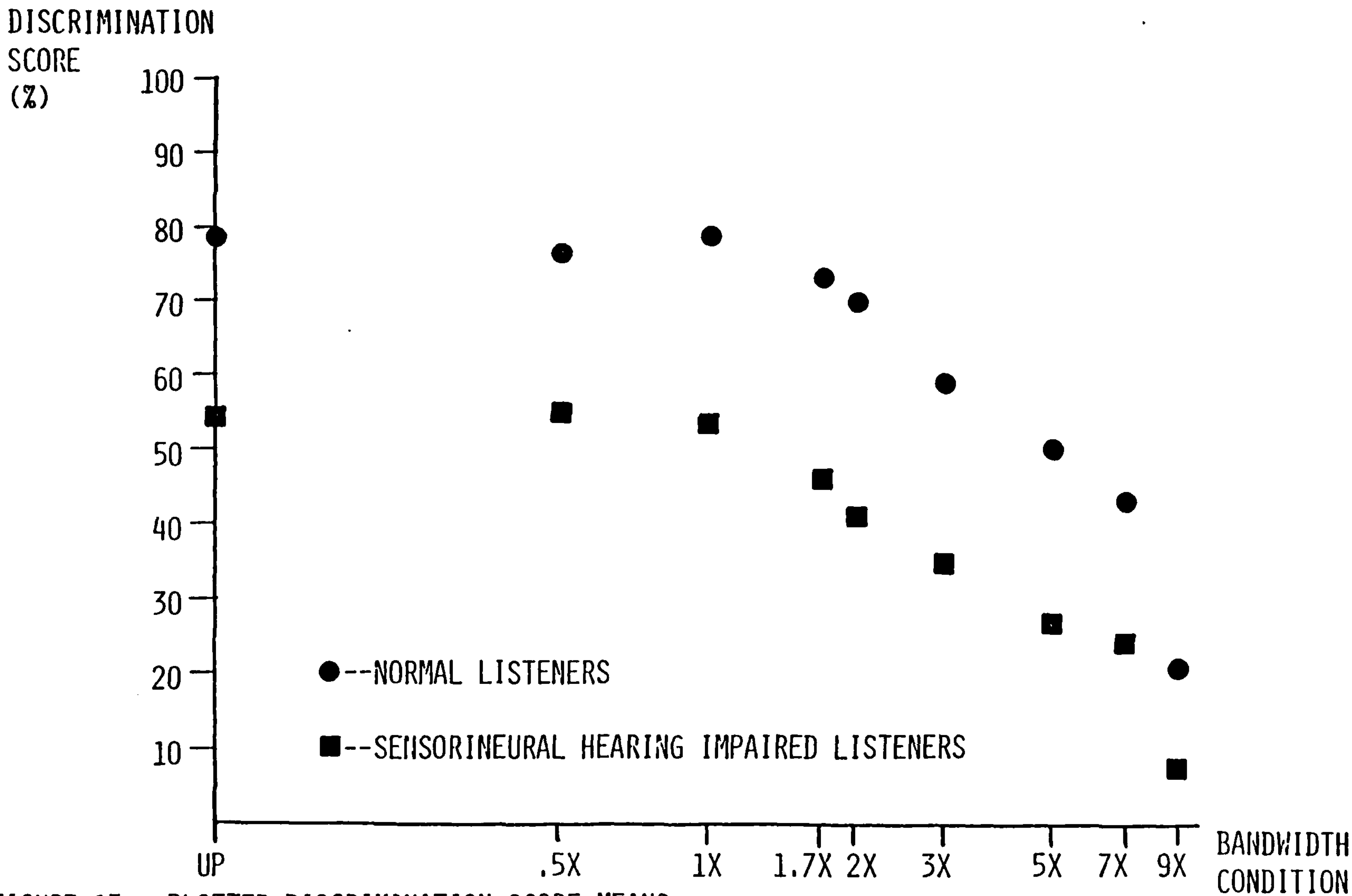

FIGURE 15. PLOTTED DISCRIMINATION SCORE MEANS.

this normal data has a slope = -36.13 and a y-intercept = 105.11. The bandwidth condition vs the sensorineural hearing impaired discrimination scores exhibited a correlation of r = -.73. The sensorineural hearing impaired group's regression line has a slope = -26.02 and a y-intercept = 81.04. The details of this best-fit computation are found in Appendix B. The normal group's average discrimination score for the UP, HX and 1X conditions is 77.56 percent; The sensorineural hearing impaired group's average discrimination score for the UP, HX, 1X and 1.7X conditions is 51.35 percent. The rationale for choosing to average these particular condition scores was derived from the results of the analysis of variance with multiple t-tests described below.

A two-factor analysis of variance with repeated measures (cf. Glass and Stanley, 1970) was performed on the discrimination data. These results are summarized in Table 5. The test indicated a significant interaction between the two factors for alpha equal to .05. However, inspection of the plotted means of the two groups (Figure 15) revealed score trends that would tend to mislead the interpretation of such a test result. The test scores for the 9X condition are disparately low in relation to the trend of the other bandwidth condition scores. The allowed bandwidth resolution available at the 9X condition may have been too extreme to allow for a fair assessment of discrimination ability. Note that the mean sensorineural hearing impaired perfor-

Table 5

Analysis of Variance Summary Table:
UP through 9X Conditions

Source	ss	ms	df	F-Ratio
Between Subjects				
A	78462.2	78462.2	1	72.64*
Error	75607.4	1080.1	70	
Within Subjects				
J	194469.7	24308.7	8	347.05*
AJ	2507.9	313.5	8	4.48*
Error	39225.1	70.0	560	

* Significant at .05 level

mance for the 9X condition was not significantly different from 0 percent. This result indicates that the allowed bandwidth resolution at the 9X condition was too rigorous to avoid limiting effects. It was decided on the basis of this result to remove these extreme data and rerun the two-factor analysis of variance test. These data are summarized in Table 6. This time the test indicated no significant interaction of the two factors for alpha equal to .05. Thus, the initial interaction predicted by the first analysis of variance test was due to the extreme distortion of the processed speech at the 9X condition, and not due to the intended effects of the signal processing.

A single factor analysis of variance for both groups indicated a significant main effect of the speech processing (see Table 5). A Newman-Keuls follow-up test (cf. Glass and Stanley, 1970) demonstrated a significant decreasing trend for both groups as the bandwidths were varied from 2X through 9X. Similarly, no significant differences in discrimination scores were observed for either group for the UP (unprocessed) through 1X conditions. Results which discriminated between the normal and sensorineural hearing impaired groups were notably those speech discrimination scores for bandwidth conditions from 1X through 2X. A priori support for multiple t-tests came from the results of the MDB test above, in which the normal group scored a 0.90X critical bandwidth rating and the sensorineural hearing impaired group scored a 1.82X critical bandwidth rating. A

Table 6

Analysis of Variance Summary Table:
UP through 7X Conditions

Source	ss	ms	df	F-Ratio
Between Subjects				
A	77388.3	77388.3	1	70.05*
Error	77330.3	1104.7	70	
Within Subjects				
J	88057.6	12579.7	7	189.14*
AJ	831.9	118.8	7	1.79
Error	32590.5	66.5	490	

* Significant at .05 level

Newman-Kuels follow-up test indicated that the normal group scores at the 1X condition were significantly larger than those at the 1.7X condition. The follow-up test on the sensorineural hearing impaired group scores indicated no significant difference between the 1X and the 1.7X scores. Neither group showed significant score differences between the 1.7X and 2X condition.

An additional two-factor analysis of variance with repeated measures was performed on the 2X through 7X portion of the discrimination data in an effort to evaluate the supra-critical bandwidth resolution performances of both groups. These results are summarized in Appendix C. The test indicated a significant interaction between the two factors for alpha equal to .05.

A series of correlation calculations were made between the unprocessed speech discrimination results, the MDB test results, and the pure-tone audiometric threshold values. A significant correlation coefficient was found between the unprocessed speech discrimination scores and the corresponding pure-tone audiometric threshold values; $r = -.56$. A significant correlation coefficient was also found between the unprocessed speech discrimination scores and the corresponding MDB test results; $r = -.34$. No significant correlation was found between the pure-tone audiometric thresholds and the MDB test results. In addition, a significant multiple correlation was found between the dependent UP discrimination scores and the independent MDB and pure-tone threshold results; $R = +.75$. Note that this multiple correlation value is larger than either of the significant two-variable correlation values.

Chapter VII

DISCUSSION AND CONCLUSIONS

As noted in Chapter I, the peripheral auditory system has been described to perform a preliminary place-specific frequency analysis of incoming acoustic signals. An efferent feedback loop pathway carries out a selective inhibition of frequency specific afferent fibers. The limit to which frequency information may be gated is called the critical band and has been observed in such psychoacoustic contexts as masking, loudness, and musical consonance. More importantly, the critical band mechanism performs noise-band limiting and harmonic discrimination, both of which are crucial for the correct perception of such complex acoustic stimuli as speech. Cochlear pathologies can affect the integrity of the critical bandwidth mechanism, which in turn can incur deficits in these functional characteristics due to widening and overlapping of the critical bands. The effect on the listener is one of resolution loss, in which his/her frequency resolving power is insufficient to enable discrimination of speech from an entire acoustic stimulus. To determine the unknown degree to which such a pathology has manifested itself, this research focused on a deductive approach, whereby the bandwidth resolution of a presented speech stimulus was controlled as the perceptual response was monitored.

To generate such bandwidth-controlled stimuli, a digital signal processing scheme was composed, taking advantage of the precise and diverse signal modification capabilities of a discrete system. Such processing algorithms are not without usage requirements and inherent processing limitations, which are reflected in the presence of the Nyquist criterion, time segment rules, and the smoothing process. Even with these requisites, the digital approach offers greater signal processing opportunities than conventional analog filters and modulators.

The use of resolution-limited tapes on subjects with normal hearing and sensorineural hearing impairment yields data by which such a processing scheme may be evaluated.

Normal Listeners

The performance of normal listeners with the processed speech test indicated three distinct trends:

> 1) A plateau effect was observed for the UP through 1X condition lists; that is, no significant differences among these scores were observed.
>
> 2) The 1X score was significantly larger than the 1.7X score.
>
> 3) A monotonic, decreasing trend was observed as the allowed bandwidth resolution was widened from the 2X through 7X conditions.

The 2X through 7X scores graphically demonstrated a close approximation to a logarithmic curve with a negative slope (Figure 15). The high correlation coefficient between these discrimination scores and their bandwidth conditions ($r = -.75$) also underscored this observation.

The MDB test confirmed that the sampled normal listeners did indeed have normal critical bands, equal to 0.90X. Note that this mean bandwidth value was statistically equal to the widest processed speech list in which non-decremental performances were observed (i.e., the 1X list). In other words, this independent critical bandwidth rating corresponds to the observed point of inflection demonstrated by the speech discrimination scores. Thus, the results for normal listeners support the notion that significant decrements in speech discrimination ability are not observed until the allowed bandwidth resolution exceeds the listener's own resolving power.

Sensorineural Hearing Impaired Listeners

The trends noted above were also observed after repeating the two tests on persons with confirmed cases of sensorineural hearing impairment. Again, the plotted speech discrimination means demonstrated a plateau/negative slope curve, this time with what appeared to be a different inflection point. The statistical analysis revealed no significant differences between the UP, HX, 1X or 1.7X condition scores, while the 2X through 7X scores demonstrated a monotonic, decreasing trend. The 2X through 7X scores also graphically demonstrated a close approximation to a logarithmic curve with a negative slope (Figure 15). The high correlation coefficient between the discrimination scores and the bandwidth condtions ($r = -.73$) emphasized this observation.

The MDB test resulted in a mean value of 1.82X, a significantly wider bandwidth than the normal critical band. Note that the sensorineural hearing impaired speech discrimination scores showed monotonic decrements as the bandwidth condition was varied from 2x through 9X, while scores less than 1.82X (i.e., the UP through 1.7X conditions) showed no significant differences. In other words, the results with sensorineural hearing impaired listeners also pointed towards a correlation between the independent critical bandwidth test and the point of inflection found on the speech discrimination curve. Thus, these data also support the notion that significant decrements in speech discrimination ability are not observed until the allowed bandwidth resolution exceeds the listener's own resolving power.

Correlation of Test Results

As noted in Chapter I, the width of the critical band has been found to be independent of the magnitude of threshold hearing loss. This finding has also been supported by the lack of significant correlation between the pure-tone audiometric threshold values and the critical bandwidths found during the MDB test.

The multiple correlation test results demonstrated a significant correlation between the dependent speech discrimination scores and the independent MDB and pure-tone threshold values. This value was higher than either of the significant two-variable correlation values. Thus, each independent measure of hearing acuity has a significant correlation to speech discrimination ability, however, use of the MDB test results in addition to the pure-tone thresholds would significantly improve estimation of unprocessed speech discrimination ability.

Limits on the Speech Discrimination Test

A priori support for inclusion of the 1.7X condition in the test regime came from the results of the critical band test. Since the sensorineural hearing impaired group's mean performance on this test equalled 1.82X, the 1.7X condition was generated to provide data that would discriminate between the sensorineural hearing impaired and normal groups. The results indicated that the total number of conditions used is sufficient for the pathologic group tested. However, in order to provide such a differential ability for sensorineural hearing impaired groups with other pathologic critical band conditions, the processing algorithm should be able to produce bandwidth-controlled stimuli at other decimal multiples of one critical band. A finite limit has been determined, however, relating to how many pathologic, non-integer processing conditions can be successfully composed and implemented. The reasons for this constraint are derived from 1) inherent features of the processing scheme, 2) certain psychoacoustic trends noted in the results, and 3) parameters of the discrimination test. Moreover, these reasons are interrelated to one another, each contributing collectively to the restriction noted above.

Consider the following inherent features of the signal processing scheme: 1) The frequency resolution of this algorithm has a finite limit. The allowed time segment size in an FFT call, which satisfies the steady-state speech assumption, combines with the maximum available sampling rate to limit the discrete spectral resolution to approximately 78 Hz/array point. 2) As implied by this resolution ratio, the values of the frequency array have a linear relationship.

The ear, on the other hand, has been observed to be a constant-percentage frequency analyzer (cf. Moore, 1977). That is, equal pitch changes are perceived with logarithmic adjustments of frequency. Fechner (1889) derived a logarithmic relationship that applies to the perception of absolute pitch magnitude. It is stated in his law:

$$S = K\log(I),$$

where: I equals the frequency magnitude, and
S represents the perceived pitch.

This phenomenon was also demonstrated by the plotted discrimination score means. When a logarithmic abscissa is employed to plot the bandwidth condition vs discrimination score curve, a nearly straight line is observed for the 2X through 7X scores for both groups. Thus, the ear's perception of such complex acoustic stimuli as speech varies as a logarithmic function of frequency bandwidth.

The accuracy of the inferential analysis performed on the discrimination data is limited by the statistics associated with the particular samples tested. For example, the value of the minimum significant difference between the discrimination scores is: 1) proportional to the subject score variability, and 2) inversely proportional to the subject sample size. Note that subject size and variability are also interdependent. If one attempts to reduce the minimum significant difference value by using an excessive sample size, the consequential increase in subject variability counteracts any advantage of this increase (Glass and Stanely, 1970).

These functional characteristics contribute to the test's design limits in the following manner: Since the discrete spectral content of the digitized signal is available as a linear function of frequency, there exists an uneven balance of available resolution ability with regards to the logarithmic receptor system of the ear. The discrete high frequency content has more information density than is necessary, whereas the discrete low frequency content is limited in its resolution by the required time segment size and available sampling rate. The center frequency of the lowest processed frequency band is 570 Hz. The critical band (i.e., the 1X condition bandwidth) at this center frequency is 120 Hz wide (Table 1). If a 1X bandwidth condition centered at 570 Hz is 120 Hz wide, then the bandwidth's percentage of the center frequency is 120/570 = .2105. Similarily, a band 78 Hz wide yields a percentage of that center frequency equal to 78/570 = .156. Since 78 Hz/array point is the minimum available spectral resolution, dividing

.156 by .2105 yields .74X; the minimum processing condition that may be generated at that center frequency. At higher frequencies, the available resolution allows for narrower bandwidth conditions. Note that the HX condition processes speech to a .5X resolution for all but the lowest frequency band, where only a .74X resolution is possible. Thus, the narrowest processed bandwidth condition that may be generated for all bands under the present algorithm is a .74X condition.

The noted statistical constraints interact with the psychoacoustic trends described above to limit the number of feasible pathologic conditions that may be generated. Using the average slope of the 2X through 7X portion of Figure 15 and the minimum significant score difference for each of the two groups, one may estimate how many other pathologic conditions are neccessary. Thus, the limits of the digital signal processing scheme, the listening trends of human listeners in response to complex frequency stimuli, and the effect of the discrimination test parameters all contribute to the finite number of processing conditions that may be generated.

It is reasonable to conclude that these factors were responsible for the non-significant difference between the 1.7X and 2X scores noted for both groups.

Implications of Plotted Discrimination Scores

The normal hearing and sensorineural hearing impaired groups both demonstrated non-decremental score trends for the narrower bandwidth conditions. As such, the normal group's plotted UP through 1X condition scores as well as the sensorineural hearing impaired group's plotted UP through 1.7X condition scores exhibited a slope equal to zero. Beyond the inflection point of each group's score trends, however, the slopes were not of the same value. The two-factor analysis of variance performed on the 2X through 7X condition scores indicated a significant interaction between the normal and sensorineural hearing impaired groups (Appendix C). These data supported the observation that the 2X through 7X portions of the two curves have unequal slopes, and hence were not parallel to each other. Furthermore, the regression curve analysis (Appendix B) indicated that the sensorineural hearing impaired group's scores exhibited a less steep slope (-26) when compared to the normal group's slope (-36).

The intercept with the UP axis was different for each curve. While the normal group's intercept value was equal to 105 percent, the sensorineural hearing impaired group's intercept value was equal to 81 percent. Hereafter, these intercept values will be referred to as target scores. A target score is hereby postulated to be the maximum speech discrimination score that can be achieved by an individual under ideal (i.e., maximum signal-to-noise ratio) listening conditions.

Note that the target score for the normal group equalled approximately 100 percent. Speech has been found to be highly resistant to distortion due to the redundancy of speech cues (cf. Rosenweig & Postman, 1957; Minifie, et al, 1973; Moore, 1977), and hence speech discrimination has been observed to follow a psychometric function (Miller, et al, 1951). In other words, normal listeners more rapidly ap- proached a maximum discrimination score of 100 percent as the level of the speech or the signal-to-noise ratio was increased (French & Steinberg, 1947). The target score for normal hearing individuals, then, appears to reflect their maximum possible speech discrimination score under ideal listening conditions.

A search of the tested sensorineural hearing impaired subjects' clinical records was made to assess this groups' speech discrimination ability under ideal listening conditions. With no added masking source, the groups' mean discrimination score with an unprocessed word list was 83 percent. This score may be thought of as an empirical measurement of the target score. The similarity between this value and the target value for the sensorineural hearing impaired group emphasized the possibility of their postulated relationship. Thus, an indication of an individual's speech discrimination ability under ideal listening conditions may be derived from the results of a speech test using supra-critical bandwidth resolution signals with a finite masking source.

It is further postulated that this target score is a constant for an individual with a given degree of hearing integrity. It would be possible to evaluate this hypothesis by testing normal and sensorineural hearing impaired groups with the UP through 7X lists at different signal-to-noise ratios. Regardless of the ratio used, both groups' scores should demonstrate a plateau (zero slope) score section for bandwidth conditions less than or equal to each group's mean bandwidth rating. The mean percent score of this plateau section is all that should fluctuate, depending on the signal-to-noise ratio used. The performance for the 2X through 7X conditions should vary with different signal-to-noise ratios in such a manner that the score's plotted curve still yields the same target (intercept) score. For example, a more stringent signal-to-noise ratio of -10 dB may cause a steeper negative slope for the conditions wider than the respective inflection points, but the mean values across all conditions should also be lower. These two factors should vary in such a manner that the target score remains the same. Therefore, as a suggestion for further research, the UP through 7X bandwidth conditions should be tested on normal and sensorineural hearing impaired individuals at signal-to-noise ratios that are more and less stringent than +10 dB, in an effort to assess what effect this variable has, if any, on the target score.

Characteristics of the Processed Speech

The reader will recall that the amplitude area of the processed speech array under pathologic conditions (i.e., 1.7X through 7X cases) is increased compared to that for the normal condition (i.e., the 1X case). This phenomenon was shown to result from the integration of energy more than once when that energy was located at a frequency which was common to more than one band in the widened critical band case. The computer interprets this effect as adding more amplitude area to a three-dimensional array. Similarly, the auditory system perceives this effect as a loudness increase. This abnormal perception of loudness increase for a given input amplitude is akin to the observed phenomenon of recruitment. Thus, this is an inherent feature arising from this processing scheme that systematically models a phenomenon that is known to occur in the sensorineural hearing impaired population.

Significance of the Present Work

The results of this research project provide a means for quantifying the critical bandwidth of persons on a clinical or pre-employment level. The complex portion of the test design (the computer generation of the test tapes) is effectively isolated from those that would administer the test. A clinically certified audiologist should be able to test subjects and interpret the results without difficulty.

This clinical test, because it is directly attuned to one common source of hearing loss (critical band widening), may prove in some cases to be far more sensitive than standard audiometrics in the early detection of hearing loss. Further, this test may prove useful as part of a pre-employment examination. Just as the distribution of normal listener thresholds includes values of less than -10 dB HL, one may expect to find individuals with narrower-than-normal critical bands. If critical bandwidth is as important a factor as is currently assumed in the perception of complex signals, it is possible that those individuals with more narrow than normal critical bands would be especially well suited to skilled listening tasks. The test, by virtue of the available test conditions, has the potential for discriminating between normal and "super" normal hearing individuals (i.e., those persons with narrower-than-normal critical bands). This ability of the test shows promise as a valuable tool for selecting potential workers whose hearing ability is especially well suited to jobs involving skilled listening such as radio and sonar operation.

In addition, a relationship has been established between the performance of listeners with a speech discrimination test and a test with tonal complexes as the stimuli. The MDB test results yielded a bandwidth resolution value that was correlated to the probable point of inflection of the speech discrimination test results. Existing literature has not suggested the possibility of such a correlation up to this point in time. That human auditory perception may be equally quantified by the use of either speech or discrete frequency stimuli reflects favorably on the theoretical assumptions upon which the tests are based.

Summary

This research has been a true interdisciplinary endeavor, drawing from such schools of thought as engineering, physiology, physics, mathematics, and psychology. In conclusion, this thesis has demonstrated:

1. the feasibility of generating speech signals of varying bandwidth resolutions less than, equal to and greater than the normal critical bandwidth using a novel digital signal processing algorithm, specifically,

    a. bandwidth-resolution limited signals such as those described in Chapter III have been produced;

    b. the processing scheme allows for variation of bandwidths from a minimum condition of .5X to integer and/or decimal multiples of the critical band.

    c. the stored array of processed speech signals demonstrates an abnormal growth of amplitude in the pathological cases (1.7X through 7X); a phenomenon akin to the recruitment seen in sensorineural hearing loss;

2. that normal listeners hearing the processed speech demonstrate three distinct performance trends, specifically,

    a. no significant differences among the UP, HX, and 1X scores were observed;

b. the 1X score was significantly larger than the 1.7X score;

c. as the allowed bandwidth resolution was systematically widened from the 2X to the 7X condition, the intelligibility scores showed a monotonic, decreasing trend;

d. the intelligibility scores for the pathologic listening conditions (1.7X through 7X) are comparable to scores obtained by a sensorineural hearing impaired listener presented with unprocessed, equivalent word lists;

3. that sensorineural hearing impaired listeners hearing the processed speech lists demonstrate performance trends that are similar yet significantly different from those of the normal groups', specifically,

a. no significant differences were noted among the UP, HX, 1X and 1.7X scores;

b. as the allowed bandwidth resolution was systematically widened from the 2X to the 7X condition, the intelligibility scores demonstrated a monotonic, decreasing trend;

4. that the critical bandwidth of the normal listeners was measured by an independent procedure, specifically,

a. that it is feasible to generate tonal complexes, such as those described in Chapter IV, that systematically widen with time from a sub-critical to supra-critical bandwidth;

b. these listeners have an average critical bandwidth of 0.90X, a value within the range of those found in other critical band tests with normal listeners;

c. this critical bandwidth rating is correlated to the discrimination test results, that is, the intelligibility scores for bandwidth conditions narrower-than- or equal-to- 0.90X are statistically equal to each other;

5. that the critical bandwidth of the sensorineural hearing impaired listeners was measured by an independent procedure, specifically,

a. these listeners have a mean critical bandwidth of 1.82X, a value statistically wider than the normal critical band;

b. this critical bandwidth rating is correlated to the discrimination test results, that is, the intelligibility scores for bandwidth conditions narrower-than- or equal-to- 1.82X are statistically equal to each other;

6. that the plotted discrimination scores exhibited different score trends for the normal and sensorineural hearing impaired groups, specifically,

a. beyond the inflection points of each

group's score trends, the slope of the sensorineural hearing impaired group's scores were not as steep as the normal group's;

b. the sensorineural hearing impaired group had a lower y- intercept value than the normal group for the 2X through 7X portion of the curve;

c. this intercept value has been labeled a 'target score' and is postulated to be the maximum speech discrimination score that can be achieved by an individual under ideal listening conditions;

d. the target score is similar to empirical measurements of normal and sensorineural hearing impaired group's speech discrimination ability under ideal listening conditions;

e. as a suggestion for further research, the UP through 7X bandwidth conditions should be tested on both groups using a variety of signal-to-noise ratios in an effort to assess what effect this variable has on the target score;

7. that the discrimination test regime has a finite number of feasible bandwidth condtions that may be generated, limited specifically by:

a. inherent features of the digital signal processing scheme,

AD-A119 049 PENNSYLVANIA STATE UNIV UNIVERSITY PARK APPLIED RESE--ETC F/G 17/2
EVALUATION OF CRITICAL BANDWIDTH USING DIGITALLY PROCESSED SPEE--ETC(U)
MAY 82 R D CELMER N00024-79-C-6043
UNCLASSIFIED ARL/PSU/TM-82-114 NL

2 OF 2
AD A
119049

END
DATE
FILMED
10-82
DTIC

b. certain psychoacoutic trends found by the results, and

c. parameters of the discrimination test;

8. that the described processing scheme is a reasonable approximation to the modeling of hearing for the purpo~~s of speech intelligibility in both the normal and pathologic cases, and that this approach warrants further study and development.

Appendix A

PURE-TONE AUDIOGRAM AVERAGES OF TESTED SUBJECTS

NORMALS

Frequency (Hz)	250	500	1000	2000	4000	8000
Threshold (db HL)	3.6	5.6	1.5	4.3	5.0	6.4
Range of Scores	-5/10	-5/10	-5/10	0/20	-5/25	-5/20

SENSORINEURALS

Frequency (Hz)	250	500	1000	2000	4000	8000
Threshold (dB HL)	22.7	30.8	40.0	55.4	65.6	68.9
Range of Scores	0/40	15/40	25/60	30/70	35/90	25/90

Appendix B

REGRESSION LINE COMPUTATION FOR 2X THROUGH 7X DATA POINTS

```
============================================================

           r = correlation coefficient
           b = slope of line
           a = ordinate intercept at UP condition
           X = log(bandwidth condition)
     Ybar(N) = mean normal group discrimination score
   Ybar(S-N) = mean sensorineural group discrimination score

     Ybar(N) =  55.7500             S.D.(N) =  11.8000
   Ybar(S-N) =  31.6800           S.D.(S-N) =   8.3600
        Xbar =   0.4254             S.D.(X) =   0.3125
   r(X/Y(N)) =   -.7480       r(X/Y(S-N)) =    -.7330

------------------------------------------------------------

        b(N) = -36.7300                a(N) = 105.1100

      b(S-N) = -25.0200              a(S-N) =  81.0400

============================================================
```

APPENDIX C

ANALYSIS OF VARIANCE SUMMARY TABLE

Source	ss	ms	df	F-Ratio
Between Subjects				
A	37120.4	37120.4	1	61.59*
Error	42187.3	602.7	70	
Within Subjects				
J	24425.3	8141.7	3	129.26*
AJ	681.3	227.1	3	3.61*
Error	13227.3	62.9	210	

* Significant at .05 level

REFERENCES

Bennett, T., and others. "Procedures for Characterizing Certain Effects of Prolonged Noise Exposure." Journal of the Acoustical Society of America, 63, Supplement 1, 1978.

Bienvenue, G., and P. Michael. "The Temporary Effects of Short Term Noise Exposure on Masking Phenomenon." Unpublished EAL Research Project, 1977.

Bienvenue, G., and others. "Digital Processing of Speech Materials in the Study of Sensorineural Hearing Imapairment." Journal of the Acoustical Society of America, 67, Supplement 1, 1980.

Bonding, P. "Critical Bandwidth in Presbycusis." Scandinavian Audiology, 8, 1979.

Capps, M. and H. Ades. "Auditory Frequency Discrimination after Transection of the Olivo-Cochlear Bundle in Squirrel Monkeys." Experimental Neurology, 21, 1968.

Castle, W. "Effects of Selective Narrow Band Filtering on Perception of PB Word Lists." Journal of the Acoustical Society of America, 36, 1964.

Chari, N. "Perception of 1/3 Octave-filtered Speech." Journal of the Acoustical Society of America, 61, 1977.

Cooley, J., P. Lewis and P. Welch. "The Finite Fourier Transform." IEEE Transaction on Audio and Electroacoustics, 17(2), 1969.

Cooley, J. and J. Tukey. "An Alogorithm for the Machine Calculation of Complex Fourier Series." Mathematical Computations, 19, 1965.

Curtis, J., Cleft Palate and Communication, eds. D. Sherman and D. Spriestersbach, (New York: Academic Press, 1968).

Dallos, P., The Auditory Periphery, (New York: Academic Press, 1973).

Desmedt, J., and P. Monaco. "Mode of Action of the Efferent Olivo-Cochlear Bundle in the Inner Ear." Nature, 192, 1961.

Dewson, J. III. "Efferent Olivo-Cochlear Bundle: Some Relationships to Stimulus Discrimination in Noise." Journal of Neurophysiology, 31, 1968.

Fechner, Gustav T., Elemente der Physchophysik, (Leipzig: Von Breitkopf & Hartel, 1889).

Fernandez, C. "The Innervation of the Cochlea." Laryngoscope, 61, 1951.

Ferraro, J. and J. Minckler. "The Human Auditory Pathways: A Quantitative Study: The Human Lateral Lemniscus and its Nuclei." Brain Language, 4, 1977.

Fex, J. "Auditory Activity in Centrifugal and Centripetal Cochlear Fibers in Cats." ACTA Physiologica Scandinavia, 55, 1962.

Fex, J. "Efferent Inhibition in the Cochlea Related to Hair Cell Deactivity: A Study of Postsynaptic Activity of the Crossed Olivo-Cochlea Fibers in the Cat." Journal of the Acoustical Society of America, 41, 1967.

Fletcher, H. "Auditory Patterns." Review of Medical Physiology, 12, 1940.

Fowler, E. "Marked Deafened Areas in Normal Ears." Transactions of the American Otological Society. 18, 1928.

Fowler, E. "Measuring the Sensation of Loudness." Archives of Otolaryngology, 26, 1937.

French, N., and J. Steinberg. "Factors Governing the Intelligibility of Speech Sounds." Journal of the Acoustical Society of America, 19, 1947.

Gelfand, S., Hearing, (New York: Marcel Dekker, 1981).

Glass, G. and J. Stanley., Statistical Methods in Education and Psychology, (Englewood Cliffs: Prentice-Hall, 1970).

Goldstein, M. "The Auditory Periphery." Medical Physiology, ed. V. Mountcastle, (Saint Louis: Mosby, 1968).

Greenwood, D. "Auditory Masking and the Critical Band." Journal of the Acoustical Society of America, 33, 1961.

Haggard, M. "Feasibility of Rapid Critical Bandwidth Measurements." Journal of the Acoustical Society of America, 55, 1974.

Kryter, K. "Speech Bandwidth Compression through Spectrum Selection." Journal of the Acoustical Society of America, 32, 1960.

Ladefoged, P., Elements of Acoustica Phonetics, (Chicago: Chicago University Press, 1962).

Lindquist, E., Design Analysis of Experiments in Psychology and Education, (Boston: Houghton Mifflin, 1953).

Lorente de No, R. "The Sensory Endings in the Cochlea." Laryngoscope, 47, 1937.

Michael, P., and G. Bienvenue. "A Procedure for the Early Detection of Noise Susceptible Individuals." American Industrial Hygine Association Journal, 1976.

Miller, G., G. Heise, and W. Lichten. "The Intelligibility of Speech as a Function of the Context of the Test Materials." Journal of Experimental Psychology, 41, 1951.

Milner, P., Physiological Psychology, (New York: Holt, Rinehart and Winston, 1970).

Minifie, F., T. Hixon, and F. Williams., Normal Aspects of Speech, Hearing and Language, (Englewood Cliffs: Prentice-Hall, 1973).

Moore, B., Introduction to the Psychology of Hearing, (Baltimore: University Park Press, 1977).

Morton, J. and A. Carpenter. "Experiments Relating to the Perception of Formants." Journal of the Acoustical Society of America, 35, 1963.

Pickles, O., and S. Comis. "Role of Centrifugal Pathways to Cochlear Nucleus in Detection of Signals in Noise." Journal of Neurophysiology, 36, 1973.

Plomp, R. "The Ear as a Frequency Analyzer." Journal of the Acoustical Society of America, 36, 1964.

Plomp, R. and W. Levelt. "Tonal Consonance and Critical Bandwidth." Journal of the Acoustical Society of America, 38, 1965.

Rabiner, L. and R. Schafer., Digital Processing of Speech Signals, ed. A. Oppenheim, (Englewood Cliffs: Prentice-Hall, 1978).

Ramon Y Cajal, S., Beitrag zur Studium der Medulla Oblongata, (Leipzig: Barth, 1896).

Rasmussen, G. "The Olivary Peduncle and Other Fiber Projections of the Superior Olive Complex." Journal of Comparative Neurology, 84, 1946.

Rasmussen, G. and W. Windle, Neural Mechanisms of the Auditory and Vestibular Systems, (Springfield, Illinois: Charles Thomas, 1960).

Remez, R. "Adaption of the Category Boundary Between Speech and Nonspeech: A Case Against Feature Detectors." Haskins Laboratories: Status Report on Speech Research SR-50, 1977.

Richards, D. and R. Archbald. "A Development of the Collard Principle of Articulation Calculation." Proceedings of the IEE, 103B, 1956.

Rosenzweig, M., and L. Postman. "Intelligibility as a Function of Frequency of Useage." Journal of Experimental Psychology, 54, 1957.

Scharf, B. "Complex Sounds and Critical Bands." Psychological Bulletin, 58, 1961.

Scharf, B. "Critical Bands." Foundations of Modern Auditory Theory, ed. J. Tobias, (New York: Academic Press, 1970).

Singleton, R. "An Algorithm for Computing the Mixed Radix Fast Fourier Transform." IEEE Transactions on Audio and Electroacoustics, 17(2), 1969.

Spoendlin, H. "The Innervation of the Cochlea Receptor." Basic Mechanisms in Hearing, ed. A. Moller, (New York: Academic Press, 1973).

Tillman, T., and R. Carhart. "An Expanded Test for Speech Discrimination Utilizing CNC Monosyllabic Words (Northwestern University Auditory Test No. 6)." Technical Report, SAM-TR-66-55, United States Air Force School of Aerospace Medicine, Aerospace Medical Division (AFSC), Brooks Air Force Base, Texas, 1966.

Trachiotis, C., and D. Elliott. "Behavioral Investigations of Some Possible Effects of Sectioning the Crossed Olivo-Cochlear Bundle." Journal of the Acoustical Society of America, 48, 1970.

Woolsey, C., and E. Walzl. "Topical Projection of Nerve Fibers from Local Regions of the Cochlea to the Cerebral Cortex of the Cat." Bulletin of Johns Hopkins Hospital, 1942.

Zwicker, E. "Die Verdecklung von Schmalbandgerauschen durch Sinustone." Acustica, 4, 1954.

Zwicker, E. "Uber Psychologische und Methodische Grundlagen der Lautheit." Acustica, 8, 1958.

Zwicker, E., and R. Feldtkeller. "Uber die Lautstarke von Gleichformigen Gerauschen." Acustica, 5, 1955.

Zwicker, E., G. Flottorp, and S. Stevens. "Critical Band Width in Loudness Summation." Journal of the Acoustical Society of America, 29, 1957.

VITA

Robert Daniel Celmer was born on November 3, 1956, in Garden City, New York. He graduated from St. Bernard's High School in Uncasville, Connecticut, in 1974. He was named a Connecticut State Scholar for the year 1974. In September 1974, he entered the University of Hartford and , in June 1978, received a B.S. degree in Mechanical Engineering from that institution. In September 1978, Mr. Celmer entered The Pennsylvania State University and, in 1980, received a M.S. in Acoustics from that institution. Since 1979, he has worked as a graduate assistant for research projects dealing with models of audition, effects on noise on man and other psychoacoustic topics at the Environmental Acoustics Laboratory of The Pennsylvania State University. Mr. Celmer is a member of the Acoustical Society of America, the Audio Engineering Society, the Institute for Noise Control Engineering, the American Society of Mechanical Engineers, Pi Tau Sigma Honorary Mechanical Engineering Fraternity, and the Honor Society of Phi Kappa Phi. He is the co-author of professional publications dealing with models of audition and has presented papers in this area at national meetings of professional associations.

DISTRIBUTION LIST FOR TM 82-114

Commander (NSEA 0342) Naval Sea Systems Command Department of the Navy Washington, DC 20362	Copies 1 and 2
Commander (NSEA 9961) Naval Sea Systems Command Department of the Navy Washington, DC 20362	Copies 3 and 4
Defense Technical Information Center 5010 Duke Street Cameron Station Alexandria, VA 22314	Copies 5 through 10

ATE

LMED

-8